

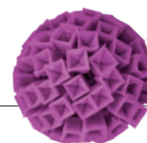
THIS WEEK

EDITORIALS

DISCOVERY Underground worms confirm that life is all around us **p.6**

WORLD VIEW China will not need anyone's help to feed its people **p.7**

CRYSTAL Hopper-shaped salts emerge from potent mixture **p.8**



A big disease with a little name

Three decades on from the first published cases of what would later be recognized as AIDS, the social and cultural challenges of the disease remain.

On the 30th anniversary of the first description of AIDS, there is much progress to celebrate, but still much work to be done. Research breakthroughs continue to improve treatments and to provide evidence for newer, better strategies that could help people to protect themselves from infection and prevent those infected from spreading the virus. Just last month, researchers reported a study, sponsored by the US National Institutes of Health, showing that when patients are treated early it reduces the chance that they will pass the virus on to uninfected partners by 96% (see go.nature.com/e5e9m7). That study is the latest, and perhaps strongest, evidence to support the concept of 'treatment as prevention' — the use of antiretroviral drugs in people who are newly infected, or even in uninfected people in high-risk groups, to slow the spread of AIDS.

In one of two Comments published in this issue (see page 29), Salim Abdool Karim at the University of KwaZulu-Natal in Durban, South Africa, details other recent successes in treatment as prevention, including a trial he conducted with colleagues in which women who used a vaginal gel containing an antiretroviral drug were less likely to become infected with HIV. Another study released last year showed that prophylactic antiretrovirals can cut the risk of new infections among men who have sex with men (R. M. Grant *et al.* *N. Engl. J. Med.* **363**, 2587–2599; 2010). As he points out: "stopping the epidemic is within our grasp".

But it is not a given that new biomedical interventions alone can slow the spread of AIDS. Six million people with HIV in developing countries now receive antiretroviral treatment, a remarkable accomplishment, but the Joint United Nations Programme on HIV/AIDS (UNAIDS) estimates that about 10 million people who need treatment aren't getting it. If the world cannot afford to treat the people who already need it, who will finance the expansion of treatment to people in the early stages of infection — as the World Health Organization recommends — or to high-risk groups who are not even infected? Advocates of universal access to antiretroviral drugs argue that providing early treatment to everyone who needs it is imperative, but how do we define 'need' when discussing preventative measures?

MONEY MATTERS

These questions take on more weight in a time of global financial trouble, when big donors such as the United States are consumed with budget cutting, and Europe's finances are being drained by bail-outs of banks and countries in financial difficulty.

In an encouraging move, US Ambassador Eric Goosby has started a discussion within the US government on what it might take for the world to treat everyone now infected with HIV, which, the study released last month suggests, would drastically slow the spread of the virus. Goosby heads the President's Emergency Plan for AIDS Relief, which already spends nearly US\$7 billion a year, and aims to treat 4 million people across 88 countries. This conversation is a helpful part of the ongoing worldwide effort by the Global Fund to Fight

AIDS, Tuberculosis and Malaria, and others to coax donor countries to underwrite the roll-out of a global strategy, based on expanded access to antiretrovirals, to drastically curtail the spread of AIDS.

Implementing such a strategy would also require overcoming a hurdle that dogs AIDS-prevention efforts. Abdool Karim writes that a staggering proportion of people do not know they are infected with HIV because of the fear and stigma that prevents them from getting tested. This situation is acute in Abdool Karim's home of South Africa, where former president Thabo Mbeki denied for years that HIV causes AIDS. But it is a problem around the world: the UNAIDS estimates that only about half of the 33 million people living with HIV know their HIV status, and this is an obvious impediment to getting treatment to everyone who needs it.

COST-BENEFIT ANALYSIS

If the cost of providing treatment for prevention seems enormous, the cost of inaction is far greater, writes Lucie Cluver at the University of Oxford, UK, in a second Comment, on page 27 of this issue.

Cluver led the developing world's first longitudinal study of the impact on children of parents made ill or killed by AIDS. The study, begun in 2005, has revealed the heartbreaking toll of the disease. Those orphaned by AIDS are more likely than children orphaned by other causes, including homicide and suicide, to develop post-traumatic stress disorder. Children of parents who are sick with AIDS are more likely to be depressed or to have an anxiety disorder than those whose parents are sick owing to other causes. Children in families affected by AIDS do worse at school and are more likely to engage in prostitution, which, in turn, makes them more vulnerable to becoming infected with HIV themselves. The toll of AIDS on the economies of developing nations can undermine these countries' attempts to escape poverty.

Poor countries now bear the brunt of AIDS, in contrast to 30 years ago, when AIDS was mainly a disease of white homosexual men in rich countries. Antiretroviral drugs are credited with curtailing the disease in these communities, but the truth is that slowing the spread required social norms to change as well.

Next week, governments will meet in New York at a United Nations meeting to plan the global response to AIDS for the next decade. They should endorse calls by UN secretary-general Ban Ki-Moon for funds to get 13 million people on antiretrovirals by 2015. The world has an ever-expanding biomedical toolbox for HIV, but to curb the disease's spread, this must be coupled with efforts to change social and cultural norms. If we do not move forwards along this path, 30 years from now the toll of the disease in terms of lives lost and human potential wasted will have multiplied. And history will harshly judge the world's inability to capitalize on its chance to avert a worsening of this disaster. ■

"If we do not move forwards, 30 years from now the toll of the disease will have multiplied."

Board the eurostar

Europe's science meeting is getting bigger and better. Now's the time to get involved for 2012.

Imagine an audience of several thousand researchers, policy-makers, educators and members of the public gathering to hear and talk about science, its applications and its impacts. Imagine discussing your work with this audience. If that idea appeals to you, keep reading, and act on what follows by the end of this month.

When a small group of researchers launched the Euroscience lobbying organization in 1997 (see www.euroscience.org), no one envisaged that its most visible manifestation would be a meeting that has steadily evolved and grown in importance over the years: the Euroscience Open Forum (ESOF). Since its launch in Stockholm in 2004, the biennial meeting has attracted a larger and more diverse range of participants, retaining a central focus on the natural sciences but with increasing attention to industry and the social sciences.

At the last meeting, in Turin in 2010, there were more than 4,000 attendees from 71 countries. According to surveys of the attendees, the 120-plus scientific sessions attracted high ratings for interest and comprehensibility and for the variety of themes. The number of journalists attending has grown to more than 300.

Europe's cities have competed to host the meeting. Next year's, on 12–16 July 2012, will be in Dublin (see www.dublinscience2012.ie).

Whatever the current financial situation, over the past decade Ireland's investment in science has grown at double the rate of the rest of its economy. This has seen the quantity and impact of the country's scientific publications grow over the past five years, and resulted in marked improvements in the international rankings of its universities. The Irish government has backed ESOF2012 energetically from the start.

Some of the themes of ESOF2012 address the frontiers of scientific knowledge. They include, for example, the power and limitations of our knowledge of the genotype with respect to our understanding of the phenotype, deep-space astronomy, quantum computation, and information science. Other themes are more philosophical: the concept of free will in cognitive science; how science provides meaning; whether e-mail is the end of history. Still others lie at the forefront of applications — 'nutriceuticals' and other innovative aspects of food, and, of course, the science and the issues surrounding energy, environment and the climate.

These themes provide a framework, but the creative spark in the ESOF meeting has always arisen in the sessions, in which researchers and others have often provoked vigorous but critically minded interest — the audiences are far from passive.

A typical session involves four panellists discussing science or technologies and the issues they give rise to. In past years, competition to run scientific sessions has been fierce, with a rejection rate for the 2010 meeting of about 50%. Encouragingly, the traffic to the proposals website shows that there is already substantial interest for the meeting next year. Readers are encouraged to submit their own proposals — instructions are at go.nature.com/wq39ya. ■

Gold mine

A trove of worms found in the deep subsurface biosphere illustrates the ubiquity of life.

Zoologists of a certain age will remember with affection *Animals Without Backbones*, a classic 1938 elementary textbook by Ralph Buchsbaum that quotes a (sadly uncredited) researcher on the ubiquity of nematodes, or roundworms. Were all the matter on Earth, it says, to be made transparent apart from nematodes — the grass, the trees, the people, the animals, even the ground beneath our feet — the shapes of all these things would still be discernible, if ghostly, from their burden of worms.

It is therefore fitting, if not hugely surprising, that nematodes have now been found in the deep subsurface biosphere, a realm hitherto believed accessible only to microbes. The nematodes recovered from gold-mine workings under South Africa — reported on page 79 of this issue — are the first multicellular organisms to be discovered in such a subterranean setting.

Although underground worms may bring to mind the imaginings of science-fiction writers, these are not monsters. The cramped spaces between the grains of rocks, where these creatures presumably graze on bacteria, have straitened them. At hardly half a millimetre in length, they would barely deserve even a slither-on role in *Tremors*, the sci-fi spectacular that stars enormous worms that destabilize buildings. And our golden nematodes would be as fleas before an elephant compared with the majestic sandworms of *Dune*, which were the size of jumbo jets.

Yet their discovery is encouraging, even heartening, for it demonstrates, once again, that once life appears, it is hard to extinguish — and that there are few, if any, habitats in which living things might not survive, even thrive, no matter how seemingly unpromising the circumstances. Even today, when you may think that we have shaken

every tree and looked under every stone on the planet, new creatures hove into view. It is rare, nowadays, to discover large vertebrates as yet unknown to science, yet such discoveries are still made.

Molecular methods have extended the reach of those formerly equipped only with traps and butterfly nets, allowing trawls of entire habitats for new forms of life — in the sea, in the upper air and even in the rich inner spaces of our guts. There is probably no slagheap too toxic, no nuclear-waste dump too radioactive, no smoking fumarole too fuming and no icy firm too frigid that it cannot be colonized by some enterprising microbe. A promise of a carbon source, a whisper of redox potential, and life will arrive. (Indeed, to encourage such endeavours, *Nature* promises a 500 g jar of Marmite to the research group that can identify the organism that digests this savoury spread. Tip: it is probably endemic in the intestines of Britons, to many of whom Marmite is an obsession, and absent in the more sensitive stomachs of Danes, who last week threatened to ban the comestible because of its evil added vitamins.)

Perhaps the habitats least explored and prospected for life are the most familiar: soil, sediments in ponds, and the seashore — the home of the meiofauna, creatures that make their living between particles of minerals and detritus. Meiofauna creatures at their largest are just at the verge of unaided visibility, down to about 40 micrometres in diameter — a world literally in the interstices, too small for every day, but too big to trouble micro- or molecular biologists. The meiofauna offer a community of tiny arthropods and a range of creatures otherwise encountered only in the dustier (and therefore most fascinating) pages of textbooks. This is the realm in which nematodes rule — and a kingdom to which we can now add, as Buchsbaum's unheralded scientist almost predicted, the ground deep, deep beneath our feet.

In some ways their world is hidden, inaccessible and unknown. In others it is in plain sight and all around us. To pull a nematode from the deep subsurface biosphere is an achievement. But we needn't look as far as that to find wonders that are new and strange — those who look most carefully will see a universe in a grain of sand. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhbnq

E. GONG



China needs no foreign help to feed itself

China can meet the demands of its growing population without importing grain from elsewhere, says Peng Gong.

China has a long-standing ambition to be self-sufficient in grain — rice, wheat, maize (corn) and soya beans — yet it imported some 95 million tonnes last year, about 17% of its domestic production. This raised concerns that its grain imports would rise, and push up the price of food. These fears were highlighted by Lester Brown of the Earth Policy Institute in Washington DC, in an article in *The Washington Post* in March, titled 'Can the United States feed China?'

I thank Lester for his warning on food security in China, but I believe it is not a matter of whether China can feed itself. It is a matter of whether the Chinese people will choose to do so.

First, some history. China's grain production quadrupled from 1950 to 2010, and last year saw the largest ever harvest. Much of the grain that China imported last year was not for consumption, but for storage in case of crises. In fact, for the past 60 years, China has, with just 7–8% of the globe's agricultural land, fed about 22% of the world's population.

China's population is projected to grow by around 9%, until it peaks at 1.46 billion around 2030. This turns the question of who will feed China into whether China can maintain self-sufficiency in grain production in the next 20 years.

There are certainly challenges. The growth of large cities should slow the pace of rural development, but China's registered-permanent-residence policy allows floating farmer workers, who spend most of their time in cities, to build houses in their home villages. Village development in the North China Plain doubled in the past decade, gobbling up 3.1 million hectares of top-quality agricultural land. And, as older farmers leave the labour force and young ones go to cities, the decline in agricultural knowledge could reduce grain yields.

Another issue is water. The north now produces more than 60% of China's grain thanks to expanded irrigation, yet groundwater supplies there continue to decrease and severe droughts have hit production. Although many global climate models predict that north China will become wetter, the region has become drier, perhaps too dry even to justify planned investment in new water infrastructure.

I am optimistic that these challenges can be overcome. China has introduced land-conversion rules that require developers and local governments to replace lost agricultural land, by draining swamp-land for example. The country has also invested heavily in improving cropland through efficient irrigation and mechanized cultivation and harvesting. And about 10 million hectares of marsh and flood land in China were converted to agricultural land between 1978 and 2008, although at a cost to ecosystems.

The government has pledged to invest four trillion renminbi (US\$600 billion) over the next ten years to secure water supplies. Properly used,

this investment could increase irrigated agricultural land by 10–20%, potentially boosting grain production by 5–10%. Even climate change could help. Glaciers in western China are likely to melt faster over the next few decades, and could water new farmland in that region. Then there is indoor, hydroponic cultivation, which has already entered China on a household scale for growing vegetables.

If per-capita grain consumption can be held stable, then to feed itself China would need only to increase grain production by 9%. Yes, the growing middle class wants to eat more meat, which requires more grain, but older people tend to eat less meat, so the demand could be balanced as the population ages. The country does not have to follow the Western model of development based on overconsumption. Thrift is deeply ingrained in the philosophy and culture of the people.

Brave policy decisions are needed. Better planning can regulate construction in rural and urban areas. Reducing the per-family area of land for residential construction would also help.

Because of the trend towards urbanization, many rural residences will become vacant in the next 20–30 years. This land must be returned to agricultural use. With this in mind, building rural structures on stilts would avoid soil destruction and enable the reclamation of built-up land. China should abolish the permanent-residence registry system and stop floating farmer workers constructing rural residences.

Furthermore, as the cold Qinghai–Tibet plateau warms, China should look to develop cropland reserves there for emergency use. And in the rural south, the government should provide financial incentives for farmers to sow farmland currently left idle by a shortage of labour.

An effective model could be worker-owned agricultural cooperatives directly linked to urban markets through subsidized transport.

Extra investment to train farmers and strengthen the role of agricultural specialists at the town and village level would protect and raise crop yields. The current promotion of college students to serve as village officers, for instance, could be enhanced with training on biodynamic farming technology and environmental protection.

With these measures, China can maintain the current level of agricultural land, and make full use of idle land. It could grow grain production, not by 9%, but by 20–30%. Given the country's remarkable achievements in population control and economic reform over the past 30 years, the world does not need to worry about having to feed China just yet. ■

IT IS NOT A MATTER
OF WHETHER
CHINA CAN
FEED ITSELF. IT IS A
MATTER OF WHETHER
THE CHINESE PEOPLE
WILL CHOOSE
TO DO SO.

➔ **NATURE.COM**
Discuss this
article online at:
go.nature.com/oyrzzr

Peng Gong is in the Center for Earth System Science at Tsinghua University, China, and the Department of Environmental Science, Policy and Management, University of California, Berkeley, USA.
e-mail: penggong@tsinghua.edu.cn

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

NEUROSCIENCE

Pay attention to the neurons

Many animals, including humans, can control the firing of certain neurons with conditioning, and researchers now show that this can lead to enhanced attention.

Robert Schafer and Tirin Moore at Stanford University in California trained two monkeys to increase and decrease the firing of neurons in an area of the brain known to be involved in eye movement, the frontal eye field. They then measured the activity at 94 sites in this region while the monkeys performed a visual search task. During the trials in which the monkeys had to increase their neuronal firing rate, their search performance improved in the part of the visual field represented by the controlled part of the frontal eye field, indicating greater attention.

The authors suggest that this may explain why neurofeedback, in which patients are trained to control their neurophysiology, might be able to treat attention deficit disorder.

Science doi:10.1126/science.1199892 (2011)

STEM CELLS

Gene repair tool for stem cells

By correcting disease-causing mutations in stem cells from patients with certain degenerative disorders, researchers could generate new disease models and treatments.

Juan Carlos Izpisua Belmonte at the Salk Institute for Biological Studies in La Jolla, California, and his colleagues engineered an adenovirus that efficiently fixes mutated portions of

the lamin A (*LMNA*) gene in stem cells created by reprogramming patients' cells. Corrected cells from patients with Hutchinson–Gilford progeria syndrome, an accelerated ageing disorder, produced normal connective tissue cells.

The same virus also repaired different *LMNA* mutations underlying a milder premature-ageing disease, atypical Werner syndrome. Moreover, it mended

LMNA mutations in adult mesenchymal stem cells, a focus of current regenerative therapies.

Cell Stem Cell doi:10.1016/j.stem.2011.04.019 (2011)

ECOLOGY

Speciation drives plant extinction

An animal's risk of extinction is associated with species traits such as body size and

geographic range — but this is not the case for plants. Jonathan Davies at McGill University in Montreal, Canada, and his colleagues report that, contrary to patterns seen in animals, threatened plants tend to be young species in rapidly evolving lineages.

The team studied the evolutionary history of threatened plant species on the Cape of South Africa — a plant biodiversity hot spot. Anthropogenic drivers, such as habitat loss, do not seem to endanger plants as they do animals. The authors suggest that conservation strategies for animals and plants should use different assessment criteria. *PloS Biol.* 9, e1000620 (2011)

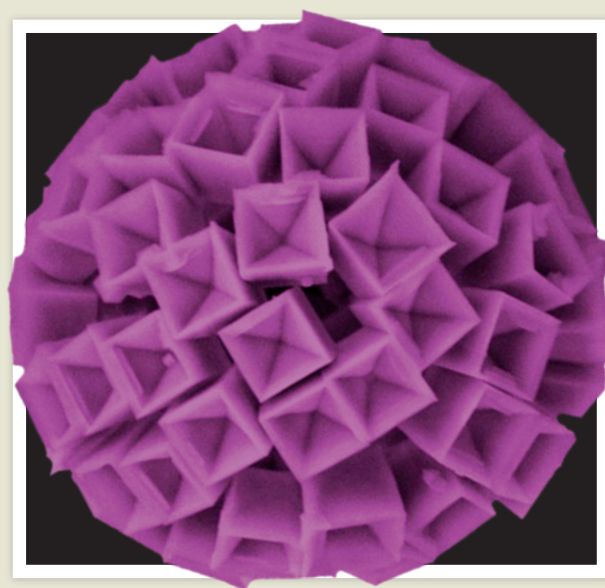
MOLECULAR NEUROSCIENCE

Bitter breath for lung airways

Lining the lower airways are 'brush' cells that can sense bitter molecules and send signals to regulate breathing, perhaps providing a defence against bacteria, which produce bitter compounds.

Gabriela Krasteva at Justus Liebig University in Giessen, Germany, and her colleagues found that brush cells in the trachea of mice express protein receptors and other key molecules involved in bitter-taste signalling. The cells make contact with vagal nerve fibres, which connect key visceral organs to the brain, and send signals to the nerves by releasing the neurotransmitter acetylcholine. When the researchers administered a bitter substance directly to the airway surfaces of mice, the animals' respiratory rate dropped suddenly.

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.1019418108 (2011)



CHEMISTRY

Crystals on water's edge

Certain salt crystals are normally cube-shaped, but can take on a new shape and grow on the surface of tiny water droplets.

Zhongping Zhang, Suhua Wang and their colleagues at the Chinese Academy of Sciences in Hefei mixed a water-based solution of either sodium chloride or potassium chloride with cyclohexane and acetone. After shaking, hollow spheres formed that were made up of individual 'hopper-like' crystals of the salt (**pictured**). The spheres probably formed at the interface between the organic solvents and the water droplets that contained the salt.

The technique may produce new crystal structures for other water-soluble compounds, and could aid in understanding crystal growth mechanisms, the authors suggest.

Angew. Chem. Int. Edn doi:10.1002/anie.201101704 (2011)



ANTHROPOLOGY

Kiwi DNA tells Maori history

The Maori people of New Zealand cherish their kiwi feather cloaks (**pictured**). An analysis of DNA extracted from the feathers offers clues to the garments' early history.

David Lambert at Griffith University in Nathan, Australia, and his team sequenced the DNA of kiwi feather samples from 109 cloaks held in museums in New Zealand and the United Kingdom. They compared the sequences to those from kiwis living in 26 locations on New Zealand's North Island.

Most of the feathers came from the North Island brown kiwi (*Apteryx mantelli*). However, 15% of the cloaks held feathers from populations in different areas, hinting at the existence of a feather trade. More than one-third of the garments included feathers from birds restricted to an eastern pocket of the North Island, which could be where cloak-making traditions started, the authors say.

Mol. Biol. Evol. doi:10.1093/molbev/msr107 (2011)

For a longer story on this research, see go.nature.com/kwyb1a

PHYSICS

Location counts in search for waves

Massive objects such as orbiting neutron stars can bend the Universe's flexible space-time fabric, generating

gravitational waves. Four wave detectors are expected to be operational on Earth by 2016. However, having an additional detector in the right place could double or even quadruple the expected detection rate, says Bernard Schutz at the Albert Einstein Institute in Potsdam, Germany.

He calculated changes in detection rate, sky coverage and directional accuracy of the network when three or more detectors are placed around the world. He found that moving one of the three detectors currently planned for the United States to Australia and adding a detector in Japan (which is now being funded) would nearly double the sky coverage. In addition, it would lower the error rate in measuring the angular position of events sixfold.

Classical Quant. Grav. doi:10.1088/0264-9381/28/12/125023 (2011)

ANIMAL BEHAVIOUR

Turtle embryos seek sunny side

Just as adult turtles warm themselves in the midday Sun, turtle embryos 'bask' in their eggs by cosy up to the Sun-warmed side.

Wei-Guo Du at the Chinese Academy of Sciences in Beijing and his colleagues shone light through the freshly laid eggs of softshell turtles (*Pelodiscus sinensis*) to mark the positions of the embryos inside. The team then incubated the eggs, warming them either from above or from one side. Over a period of days, embryos made

COMMUNITY CHOICE

The most viewed papers in science

MOLECULAR BIOLOGY

Cells stressed from hunger

HIGHLY READ
on www.cell.com
24 April–24 May

Cells that divide abnormally are starved of the building blocks needed to make DNA, placing stress on DNA replication, introducing mutations and destabilizing chromosomes. These could be some of the earliest events to occur in cancer, which is marked by widespread DNA damage.

Batsheva Kerem at the Hebrew University in Jerusalem, Israel, and her colleagues found that activating a cancer-promoting pathway in human cells grown in culture decreased cellular levels of nucleotides — the chemical building blocks of DNA. Adding the building blocks to the cells reduced DNA damage and decreased the frequency with which cells became cancerous. DNA damage also lessened when the researchers increased the transcription of genes involved in nucleotide biosynthesis.

Cell 145, 435–446 (2011)

their way to the warmest part of the egg (**pictured**: heat from above, left; from the left, right), which could differ from the coolest part by almost 1 °C. If the heat source shifted, the embryos followed.

Greater exposure of the embryos to heat could modify traits such as body size and sex, and hasten hatching to reduce nest predation.

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.1102965108 (2011)

MICROBIOLOGY

Protein behind protozoan power

Humans can become infected with the protozoan *Toxoplasma gondii* by ingesting it during its

spore-like stages, which are shed in cat faeces. The three main strains of *T. gondii* vary widely in terms of virulence in mice, and researchers have pinpointed a gene that may underlie this difference.

David Sibley at Washington University in St Louis, Missouri, and his co-workers generated a genetic cross between the type I strain, which is lethal to lab mice, and the milder type II. The progeny were injected into mice to assess their virulence. A genome-wide association analysis of the progeny and the parental strains revealed a locus on chromosome XII that accounted for 90% of the heightened virulence. Deleting a gene within this region, *ROP5*, in the type I strain completely blocked acute virulence.

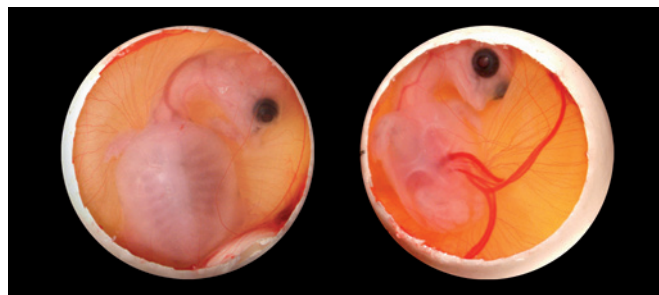
The authors suggest that the *ROP5* protein may regulate other proteins that control virulence in the parasite.

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.1015338108 (2011)

➔ **NATURE.COM**

For the latest research published by Nature visit:

www.nature.com/latestresearch



SEVEN DAYS

The news in brief

POLICY

South Africa budget

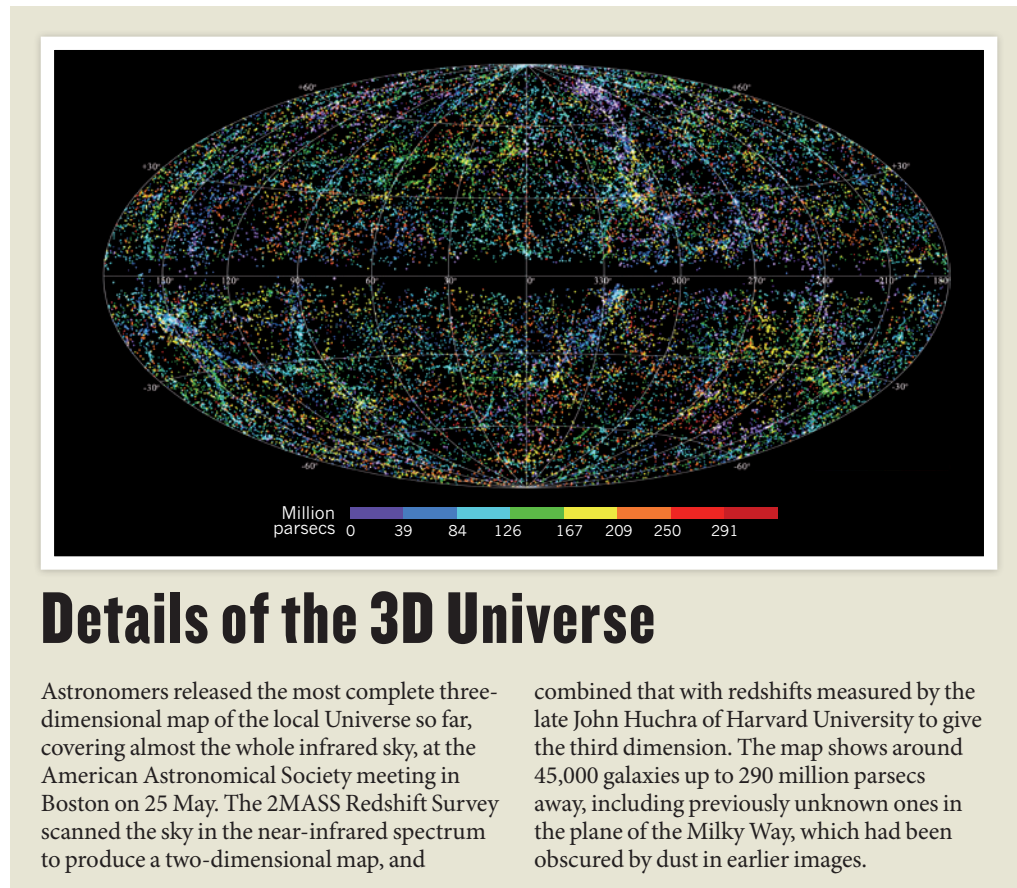
South Africa's budget for public research and development gained a 7% annual increase to 4.4 billion rand (US\$628 million) in 2011–12. Speaking after the budget announcement on 24 May, the nation's science minister, Naledi Pandor, argued that all science funding should be brought into her own portfolio, rather than spread among several departments. She also announced that South Africa would get a national committee on science, technology and innovation, whose members would include politicians and scientists. See go.nature.com/tzraqm for more.

Scientists on trial

Six Italian seismologists and one government official will be tried for the manslaughter of some of those who died in the earthquake that struck the city of L'Aquila on 6 April 2009. The city's public prosecutor had argued that the scientists falsely reassured the public. Last week a judge decided that the trial should proceed, and it will begin on 20 September. See page 15 for more.

Nuclear stress tests

The European Union is pushing for nuclear reactors worldwide to undergo stress tests like those agreed for Europe's 143 nuclear reactors. On 25 May, European regulators announced criteria for the safety reviews, which will judge how well reactors can withstand natural disasters, power blackouts and accidental explosions. The criteria do not include discussion of preparedness for terrorist attacks, which will be assessed separately — and probably confidentially. The



T.H. JARRETT/IPAC/SSC

Details of the 3D Universe

Astronomers released the most complete three-dimensional map of the local Universe so far, covering almost the whole infrared sky, at the American Astronomical Society meeting in Boston on 25 May. The 2MASS Redshift Survey scanned the sky in the near-infrared spectrum to produce a two-dimensional map, and

combined that with redshifts measured by the late John Huchra of Harvard University to give the third dimension. The map shows around 45,000 galaxies up to 290 million parsecs away, including previously unknown ones in the plane of the Milky Way, which had been obscured by dust in earlier images.

European Commission hopes that reactors which fail the tests will be upgraded or closed; it will present a preliminary report in December. Separately, Germany's government said on 30 May that it plans to shut down all 17 of its nuclear power stations by 2022; five days earlier, Switzerland's government said it would close its five nuclear reactors over the next two decades. See go.nature.com/31nwy1 for more.

Carbon high

In 2010, anthropogenic carbon dioxide emissions reached a record 30.6 gigatonnes, according to estimates released by the International Energy Agency (IEA) in Paris on 30 May. The total was 5% more than the previous record, set in 2008, says the IEA. In 2009,

emissions had dipped slightly below 29 gigatonnes as a result of the global economic crisis. See go.nature.com/rtgd7f for more.

EVENTS

E. coli outbreak

Germany is experiencing one of the world's largest ever outbreaks of enterohaemorrhagic *Escherichia coli*. By 31 May, there were 1,400 confirmed or suspected infections, 373 cases of haemolytic-uraemic syndrome (HUS — a complication that can cause kidney failure) and 14 deaths. The bacterial strain responsible — O104:H4 — has never before been associated with an infection outbreak, and its source has not been confirmed. An experimental

and expensive monoclonal antibody, eculizumab, has been pressed into service to treat some HUS patients.

RESEARCH

Asteroid mission

NASA is planning a 2016 launch for its first spacecraft aimed at returning material from an asteroid, officials announced on 25 May. The Origins–Spectral Interpretation–Resource Identification–Security–Regolith Explorer (OSIRIS–REx) will visit the carbonaceous asteroid 1999 RQ36, and NASA hopes that it will collect more material than was gathered by Japan's pioneering Hayabusa mission, which in June 2010 returned around 1,000 grains from the Itokawa asteroid. The mission

will cost US\$800 million, or \$1 billion including the cost of a launch vehicle. It beat two competing mission concepts: returning a sample from the dark side of the Moon, and a visit to Venus. See go.nature.com/2fup1u for more.

Chimp research

A US Institute of Medicine committee charged with determining whether “chimpanzees are or will be necessary for research discoveries” opened its investigation on 26 May. The National Institutes of Health (NIH) requested the study in January, after three US senators wrote to the agency on the subject. They were reacting to a controversial NIH proposal to move 186 government-owned chimpanzees out of semi-retirement and back into use in active research. See go.nature.com/m54pgh for more.

Farewell, Spirit

NASA has finally called time on attempts to contact its Mars rover Spirit (pictured). No communications had been received from the rover since March last year, when it was known to be trapped in a sandpit. After its 2004 landing on the planet, Spirit lasted much longer than the three months originally intended (see *Nature* **463**, 600; 2010). Spirit's fellow rover, Opportunity, is still exploring



Mars, and NASA's next Mars mission — the rover Curiosity — is set to launch late this year.

Well-being index

The Organisation for Economic Co-operation and Development (OECD) has launched its first index to compare well-being across countries. The index is part of the organization's ‘Better Life’ initiative — its contribution to a movement that hopes to abandon gross domestic product as the overriding measure of social development and living standards (see *Nature* **463**, 849–850; 2010). It uses 21 indicators in 11 topics, but leaves the user to assign custom weightings to these topics when comparing countries.

Rid of rinderpest

The devastating cattle disease rinderpest has been eradicated from the globe. The World Organisation for Animal Health made the formal proclamation on 25 May at its

general meeting in Paris. A global effort using widespread vaccination to eradicate the virus that causes the disease was launched in 1994. The last remaining pockets of the virus are thought to have been wiped out by 2007, although scientists have had to wait for confirmation from further cattle surveillance (see *Nature* **462**, 709; 2009).

Animal lab warning

The laboratory responsible for the UK outbreak of foot-and-mouth disease four years ago has been reprimanded again for failing to meet safety standards. It emerged last week that the Health and Safety Executive told the Institute for Animal Health at Pirbright in Surrey to improve practices after the lab reported incidents requiring investigation earlier this year. In one case, a sample of the virus that causes foot-and-mouth was being held in a cracked flask; in another, some waste liquid leaked in an incinerator room. See go.nature.com/wabyka for more.

Brain campaign

An initiative that brings together leading neuroscientists, politicians, health-care companies and celebrities aims to raise large amounts of money for brain research. ‘One Mind for Research’ was launched

COMING UP

5–9 JUNE

Canada's Perimeter Institute for Theoretical Physics in Waterloo, Ontario, hosts an ‘Equinox Summit’ to discuss how the world's energy supply should look in 2030.

www.wgsi.org

6–11 JUNE

Results from the Large Hadron Collider are presented at the ‘Physics at LHC 2011’ congress in Perugia, Italy.

www.pg.infn.it/plhc2011

9 JUNE

NASA plans to launch its Aquarius satellite, which will monitor the saltiness of the oceans.

<http://aquarius.nasa.gov>

by former Democratic congressman Patrick Kennedy — nephew of former US president John F. Kennedy — at a meeting in Boston, Massachusetts, last week. With lofty goals, the effort hopes to convince funders to reduce their focus on specific neuropsychiatric diseases, and instead support basic research on the fundamental workings of the brain. Early ambitions include a vigorous fund-raising campaign. See go.nature.com/jmytbl for more.

BUSINESS

Quantum sale

On 25 May, D-Wave Systems of Burnaby, British Columbia, Canada, announced the world's first sale of a commercial quantum computer. The purchaser was global security firm Lockheed Martin, headquartered in Bethesda, Maryland. See page 18 for more.

► NATURE.COM

For daily news updates see:

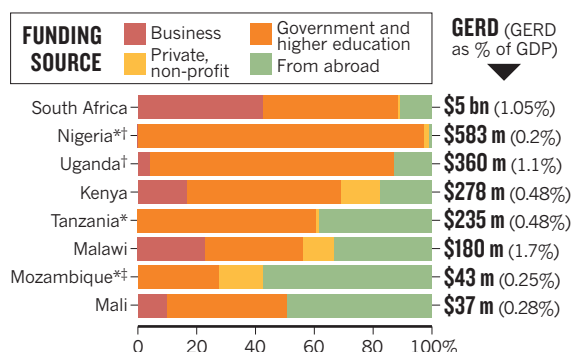
www.nature.com/news

TREND WATCH

The first comprehensive African-led initiative to survey science spending on the continent showed that South Africa, Uganda and Malawi invested more than 1% of their gross domestic product (GDP) on science and technology in 2007. But countries spending least in absolute terms tend to be more reliant on foreign funding (see chart). The survey, with detailed data from 13 countries, was unveiled at a conference in Addis Ababa last week. See go.nature.com/hmo3b5 for more.

AFRICA'S RESEARCH SPENDING

A comprehensive survey flags up some countries' dependence on foreign investment for research.



*Data do not include business enterprise sector; †data do not include private non-profit institutions/organizations; ‡data do not include the higher-education sector. GERD: gross domestic expenditure on research and development.

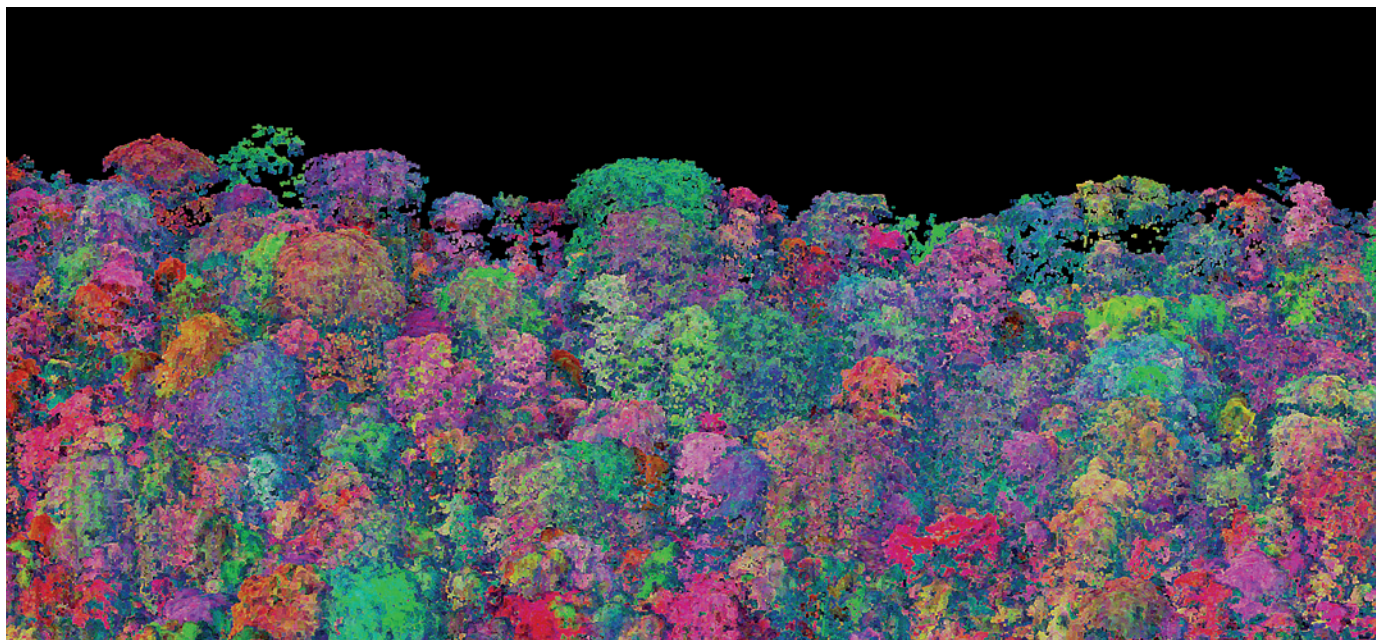
NEWS IN FOCUS

ITALY Seismologists face manslaughter charges over earthquake deaths **p.15**

ENERGY Japan disaster deals another blow to ITER fusion project **p.15**

TECHNOLOGY Quantum computers hit the market **p.18**

PROFILE Paul Davies asks cosmic questions — about cancer **p.20**



A three-dimensional image of a forest in Panama, based on data from the Carnegie Airborne Observatory. The retooled observatory will be far more sensitive.

ECOLOGY

A new eye on biodiversity

Airborne observatory will use chemical clues to map and assess tropical ecosystems.

BY JEFF TOLLEFSON

For tropical ecologist Greg Asner, it's all about seeing the forest through its trees. Over the past two years, he and his team at the Carnegie Institution for Science in Stanford, California, have used world-class tree climbers, bows and arrows, and even shotguns to gather samples of vegetation from forest canopies around the globe. They have created a digital catalogue of the chemical and optical properties of some 4,700 plant species in different conditions. Now comes the moment of truth.

On 2 June, Asner and his team will unveil the latest version of the Carnegie Airborne Observatory (CAO), an aircraft that will combine a state-of-the-art optical sensor with a laser capable of mapping forests in unprecedented three-dimensional detail. The system will allow Asner to build on earlier work cataloguing forest

carbon stocks in support of efforts to reduce deforestation (see "Taking stock of global carbon"), and will significantly advance the team's biodiversity research. With the digital catalogue as a reference, Asner hopes that the observatory will be able to perceive the species of many individual trees by their optical properties, while offering insights into forest health and diversity.

The team's work combines physics, biochemistry and ecology, beginning with measuring subtle differences in the way the forest canopy absorbs and reflects solar radiation. The signal varies depending on the leaves' concentrations of nutrients, minerals, pigments such as chlorophyll, and the compounds that plants use

➔ **NATURE.COM**
For more on Greg Asner's work in tropical forests, see: go.nature.com/75goqn

to protect themselves against the Sun and predators. In specimens from one region of the Amazon rainforest in southern Peru, Asner and his

wife, Robin Martin, identified 21 spectral traits that provided identifying signals for 90% of the species. "A lot of people look at trees and just see green," says Asner. "I see a kaleidoscope."

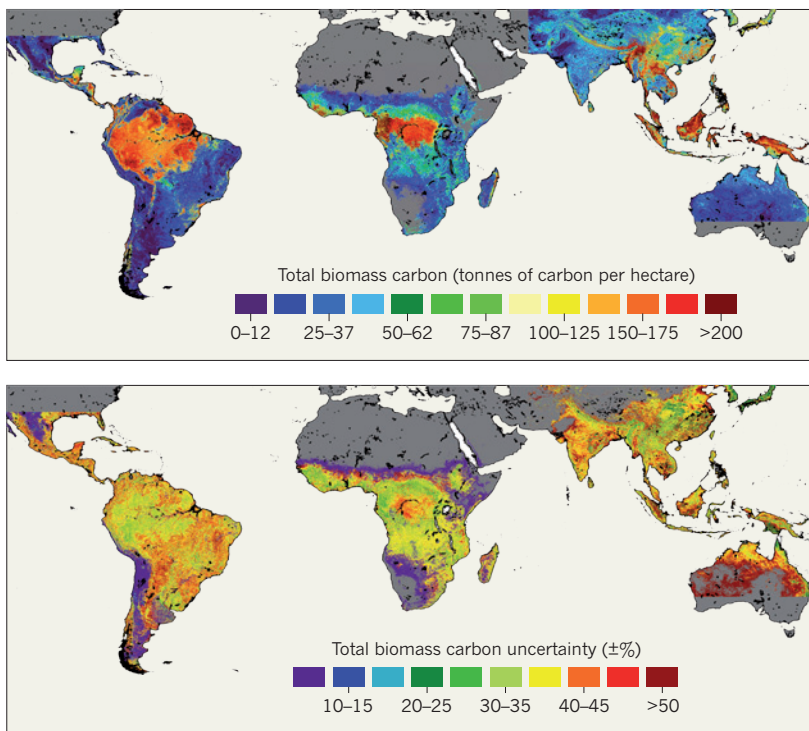
The heart of the CAO's US\$8.3-million sensing system — dubbed the Airborne Taxonomic Mapping System (AToMS) — is a spectroscopic imager designed by engineers at NASA's Jet Propulsion Laboratory (JPL) in Pasadena, California. Capable of registering more than 400 frequencies of light, from ultraviolet to infrared, the instrument will take 60,000 measurements per second, with great accuracy. "This really has taken everything we have learned at NASA and brought it to bear in the most advanced airborne imaging spectrometer ever built," says Rob Green, lead engineer on the project at the JPL.

With data from a single flight that used an earlier version of the system, Asner and his colleagues identified more than 30 species. ►

CAO/CARNEGIE INST. FOR SCIENCE

BENCHMARK OF BIOMASS

Map showing the estimated biomass of carbon (top) and the uncertainty in those estimates (bottom).

**REMOTE SENSING***Taking stock of global carbon*

This week, researchers unveiled what they hope will be a benchmark assessment of the carbon stored in the world's tropical forests. The map (pictured), which spans 75 developing countries across tropical regions, shows not only the amount of carbon stored in forests as of the early 2000s, but also the accuracy of the estimate.

"It's more than a beautiful picture," says Sassan Saatchi, a remote-sensing specialist at NASA's Jet Propulsion Laboratory in Pasadena, California, who led the research. "This is a benchmark that can be used as a basis for comparison in the future."

Forests are normally carbon 'sinks' that take up and store carbon, but logging and natural disturbances can turn them into temporary or permanent carbon sources. Knowing how much they hold and release is crucial for international negotiations that seek to pay developing countries to keep forests standing. But efforts to assess global carbon stocks and how they are changing have been hampered by deficiencies in remote-sensing technologies and a lack of data confirming satellite observations on the ground. Now

Saatchi and his team have developed a methodologically consistent way to combine the available data into a global estimate of carbon stocks.

Focusing on forests in the Southern Hemisphere, in tropical parts of Asia and in central America, the team combined laser-based measurements of forest height from NASA's ICESat mission with other remote-sensing data, as well as with information from more than 4,000 ground-level inventories of tree height and biomass. This enabled the team to map forest structure and carbon density down to a resolution of 1 kilometre. The 2.5-billion-hectare study area holds a total of 247 gigatonnes of carbon, the team estimates (S. Saatchi *et al. Proc. Natl Acad. Sci USA* doi:10.1073/pnas.1019576108; 2011).

"To have efficient negotiations on the forest contribution in reducing greenhouse-gas emissions we need improved data," says Sten Nilsson, a forest scientist with the Swedish company Forest Sector Insights and a researcher at the International Institute for Applied Systems Analysis in Laxenburg, Austria. "This work is one contribution in this direction." **Quirin Schiermeier**

► The latest technology, he says, will be three to six times more sensitive, and should enable the detection of many more plants.

No one can say for certain what the implications of these new data might be, but ecologists — many of whom have spent decades tramping through the jungle in muddy boots — are lining up to find out. "It is going to change a variety of fields of tropical ecology," says Alan Townsend, an ecologist at the University of Colorado at Boulder. "How much and in what ways we don't really know, but there is no doubt that we are starting to walk down a new path."

Peter Vitousek, an ecologist at Stanford University in California, says that he started off with the same mindset as many of his colleagues: suspicion of yet another over-hyped remote-sensing technology. On his first collaboration with Asner on a project in Hawaii several years ago, however, Vitousek saw how remote-sensing data could be used to identify invasive species and to tease out relationships among plants. After the aerial data identified areas in which the canopy was unusually depleted of nitrogen, the team went in on the ground and discovered that an invasive ginger plant was competing with native trees for the nutrient. Here was a case in which the spectral data themselves led the researchers to discover an ecological interaction, rather than simply confirming data gathered on the ground, says Vitousek. "I needed to be persuaded that this was really something new, and that persuaded me."

Asner is set to fly the CAO to the Amazon later this month, with a dual research focus. One goal is to survey the rainforest to gauge the impact of last year's drought, which seems to have been even more severe and widespread than that of 2005 — the worst event of its kind for 100 years, and one that scientists are still studying. The other is to fly over plots monitored by the Amazon Forest Inventory Network (RAINFOR) to address a nagging question: how representative are the areas that scientists have been studying for decades?

RAINFOR tracks long-term data on soil and plant growth, but Asner's information will provide unprecedented detail on the chemistry of leaves in the canopy. The ground plots provide a record of how each tree in the area fared through the drought, and Asner's crew should be able to pick those trees out of their broader analysis to look for any lingering effects. "The hypothesis is that those trees that have suffered during the drought also have different canopy chemistry," says Oliver Phillips, an ecologist at the University of Leeds, UK, who coordinates RAINFOR. "It's very exciting," he says. "But we'll see how it pans out."

Asner also has his own agenda, which begins with mapping out tropical biodiversity. "One thing is for sure: the array of chemical traits in these upper canopies is amazing and radically under-appreciated," he says. "We are going to use that to our advantage to get canopy diversity mapped." ■

SEISMOLOGY

Scientists on trial over L'Aquila deaths

Seismologists charged for giving apparent reassurances on Italian earthquake risks.

BY NICOLA NOSENGO IN ROME

The perils of communicating scientific uncertainty when under the media spotlight are set to be probed in an Italian court later this year. The case, which was given the go-ahead by a judge last week, involves six Italian seismologists and one government official. They will be tried this autumn for the manslaughter of some of the 309 people who died in the earthquake that struck the city of L'Aquila on 6 April 2009. If convicted, they could face jail sentences of up to 12 years.

The seven were on a committee tasked with assessing the risks of increased seismic activity in the area. At a press conference following a committee meeting a week before the earthquake, some members assured the public that they were in no danger. After the quake, many of the victims' relatives said that because of these reassurances they did not take precautionary

measures, such as leaving their homes.

L'Aquila's public prosecutor, Fabio Picuti, argued last week that although the committee members could not have predicted the earthquake, they had translated their scientific uncertainty into an overly optimistic message. The prosecution has focused on a statement made at the press conference by accused committee member Bernardo De Bernardinis, who was then deputy technical head of Italy's Civil Protection Agency. "The scientific community tells me there is no danger," he said at the time, "because there is an ongoing discharge of energy. The situation looks favourable."

Many seismologists — including one of the accused, Enzo Boschi, president of the National Institute of Geophysics and Vulcanology in Rome — have since criticized the statement as scientifically unfounded. The statement does not appear in the minutes of the committee meeting itself, and the accused seismologists

say they cannot be blamed for it. De Bernardinis's advocate insists that his client merely summarized what the scientists had told him. The prosecutor claims that because none of the other committee members immediately corrected De Bernardinis, they are all equally culpable.

Boschi says that he is "devastated" by the ruling. He notes that there are hundreds of seismic shocks every year in Italy: "If we were to alert the population every time, we would probably be indicted for unjustified alarm," he said, adding that poor building standards were the main cause of the tragedy.

Vincenzo Vittorini, a physician in L'Aquila whose wife and daughter were killed in the earthquake and who is president of the local victims' association, hopes the trial will lead to a thorough investigation into what went wrong. "Nobody here wants to put science in the dock," he says. "All we wanted was clearer information on risks in order to make our choices". ■

ENERGY

Japan quake rocks fusion project

Damaged facilities force further delay to ITER experiment.

BY GEOFF BRUMFIEL

IN ST-PAUL-LEZ-DURANCE, FRANCE

The world's largest fusion experiment is finally beginning to take shape. Workers at a vast site in southern France have dug the 17-metre-deep pit that will house the ITER reactor, and will soon install 500 pillars of steel-reinforced concrete that should protect the machine during an earthquake. But even as they toil, a quake halfway around the world has struck a blow to the project.

The 11 March earthquake and tsunami that hit Japan, one of seven partners in ITER, severely damaged key facilities for testing the reactor's components. Unless repairs can be made or work reassigned quickly, the damage could cause a delay of "perhaps several years", according to Osamu Motojima, ITER's director. Motojima says that he and his team are looking at ways to reduce the impact. "At present my target is less than one year's delay," he says.

ITER's first experiments have already been pushed back from 2016 to 2019, and the project has suffered serious cost overruns since its partners agreed to go ahead in 2006. Any extra delays are likely to increase political pressure to find cost savings and speed up work.

The giant reactor is designed to prove that useful energy can be extracted from the fusion of hydrogen isotopes. By trapping the hydrogen using powerful superconducting magnets and heating it to the point of fusion at 150 million °C with specially designed ►



Construction of the ITER fusion reactor in France is beset by financial and technical problems.

ALTIVE

► heating systems, ITER is supposed to produce ten times more energy than it consumes. The Naka Fusion Institute of the Japan Atomic Energy Agency, which is about 100 kilometres north-east of Tokyo, was a key facility for testing and developing the magnets and heating systems. But it was hit hard by the quake. “The buildings are damaged and we can’t go inside,” says Hiroshi Kataoka, director of the fusion division at Japan’s Ministry of Education, Culture, Sports, Science and Technology, which oversees Japan’s contributions to the ITER project. “We are not sure, but it may have some impact on our ability to deliver products on time.”

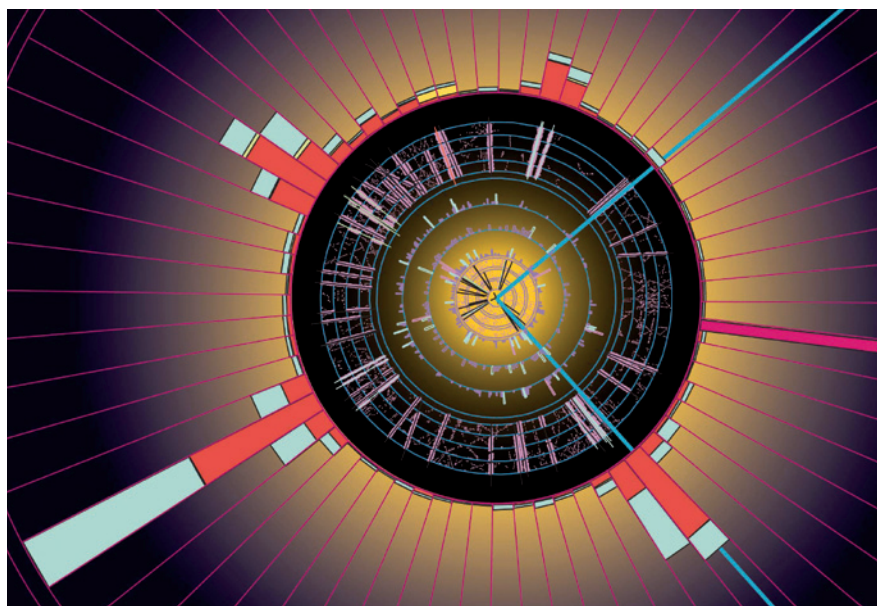
The magnet test facilities are particularly important for the project. Late last year, a sample of superconducting cable to be used in a central magnet failed testing at a facility in Switzerland (see *Nature* 471, 150; 2011). Follow-up tests at the Naka institute were expected to help determine the cause of the failure and aid any redesign of the cable, which is being manufactured in Japan. The cable factories themselves seem to be undamaged, according to Richard Hawryluk, one of ITER’s deputy directors-general.

Motojima says he is reluctant to shift work away from Japan unless absolutely necessary, but if the facilities cannot be repaired within six months, he says that he will seek arrangements with other partners in the project. A new schedule and any adjustments to the reactor’s work plan must be ready by December, he says.

Meanwhile, the project’s financial woes continue. Since 2006, ITER’s construction costs have roughly tripled to around €15 billion (US\$21 billion). The European Union, which is paying for some 45% of the project, has yet to find the additional €1.3 billion it needs to meet its near-term commitment in 2012 and 2013. Anne Jensen, a member of the European parliament from Denmark who sits on the parliament’s budget committee, says that she and other members are generally supportive of the project, as long as it does not draw money away from other areas of research. “We think ITER is an interesting project, but it should not be at the expense of the development of wind energy or smart grids,” she says. But she adds that she is optimistic that ITER’s funding can be found by the 2012 deadline.

The other partners — Russia, South Korea, India, China and the United States, which along with Japan are each contributing around 9% to the construction — are also struggling to come up with additional funding. In the United States, for example, battles in Congress over discretionary spending have meant that ITER got its 2011 allocation of \$80 million only last month.

Given the variety of financial and technical difficulties facing ITER, a further delay of “a couple of years” is probably inevitable, says Stephen Dean, president of advocacy organization Fusion Power Associates in Gaithersburg, Maryland. But, he adds, “that doesn’t mean that they can’t get it done.” ■



FERMILAB

Data from Fermilab, such as this reconstruction of two top quarks, could be lost to physicists forever.

PARTICLE PHYSICS

Tevatron’s legacy set to disappear

Lack of long-term preservation plan threatens to leave key information inaccessible for future analysis.

BY EUGENIE SAMUEL REICH

Four months before the Tevatron shuts down for good, physicists at Fermilab’s giant particle collider near Batavia, Illinois, are pulling out all the stops to collect every last bit of data that they can. But some worry about what will eventually happen to the trove of data — approaching 20 petabytes (20×10^{15} bytes) — amassed over the machine’s 26-year life. Although there is funding to continue sifting the data for traces of the Higgs boson and other subatomic prizes for the next five years, so far there is no plan and no budget for preserving them in the longer term.

At a workshop on data preservation at Fermilab on 16–18 May, some physicists called for that to change, arguing that Tevatron data could prove useful as an independent check on its successor, the Large Hadron Collider (LHC) now operating at CERN, Europe’s particle-physics lab near Geneva, Switzerland. If researchers suspect that the LHC has spotted new physics, particularly at the lower end of its energy range, the claim could be tested for consistency with Tevatron

data, says Rob Roser, spokesman for the Collider Detector at Fermilab (CDF), one of the Tevatron’s two principal experiments.

Although many fields of science, from genomics to astrophysics, put substantial resources into archiving data and making them publicly available, the norm in particle physics has, until recently, been very different. When the analysis of data from an experiment trickles to a halt, researchers typically move on. The data languish or are even destroyed to make storage space available for something else. When the Tevatron was built, “we did not think about data preservation”, says Qizhong Li, computing coordinator for D0, the other main experiment at the Tevatron. “This is a rather new concept.”

Both D0 and the CDF expect to lose their dedicated computing infrastructure over the next five years. A gradual loss of knowledge about how to deal with the complex data, which includes raw detector readouts, reconstructed particle trajectories and higher-level analyses, could also present a serious

► **NATURE.COM**
For more on how to cope with big data sets, see:
go.nature.com/vnklyt

hurdle to exploiting the data in the future.

Such neglect would be a mistake, says Cristinel Diaconu of the Centre for Particle Physics in Marseilles, France, who leads the H1 experiment at the Hadron–Electron Ring Accelerator (HERA) in Hamburg, Germany, which closed in 2007. “We always have new ideas that can be used to reanalyse data,” he says. Diaconu, who is chair of an international study group on data preservation and long-term analysis in high-energy physics, estimates that good data preservation can increase the scientific potential of an experiment by 10% for less than a 1% increase in cost.

At BaBar, an experiment that produced B mesons at the Stanford Linear Accelerator Center in Palo Alto, California, until 2008, physicists have begun building a US\$500,000 archival system that will save the raw data and software, and are also setting up virtual interfaces to run the older software on modern machines. “We have decided to save everything at least to 2018,” says Tina Cartaro, BaBar’s computing coordinator.

Other experiments are adopting cost-saving compromises. Collaborators on the HERA H1 experiment have decided that it is not necessary to keep all their seven attempts at reconstructing particle events from the raw data. “Our initial thinking was to keep everything, but we now think we will keep three iterations,” says David South, computing coordinator for H1 at the Technical University of Dortmund in Germany.

The Tevatron can learn from those examples, says Roser. But Li says it’s a tougher job than it would have been if data preservation had been planned for from the beginning. In addition, over the next five years, thousands of tapes’ worth of data will somehow have to be migrated to newer, higher-density storage and a suitable retrieval system to go with it.

The situation at the LHC is strikingly different. Computing specialists there are already working towards permanent archiving of the data, says Elizabeth Sexton-Kennedy of Fermilab, who works on computing systems for the CMS, one of the LHC experiments. In its short lifetime, the LHC has collected five times as much data as the Tevatron. All the raw data are being kept, although the CMS is saving space by deleting old attempts at reconstruction when they are surpassed by newer ones. “When we know things much better, we delete older knowledge,” says Sexton-Kennedy. “It’s the tension between the old and the new.”

Particle physicist Siegfried Bethke of the Max Planck Institute for Physics in Munich, Germany, who spent two years reconstructing data that hadn’t been maintained from PETRA, a positron–electron collider that ran from 1979 to 1986 at the DESY accelerator in Hamburg, told the data-preservation workshop that his experience indicates that better planning is really vital. “These data have cost a lot of money to the taxpayer and not conserving them would be a crime,” he said. ■

EDUCATION

Furore erupts over Turkish test cipher

University entrance-exam fiasco hits election campaign.

BY ALISON ABBOTT

Imagine sitting your university entrance examinations, only to learn the next day that the correct answers to many of the multiple-choice questions were signalled by a secret code in the exam paper itself.

For 1.7 million Turkish high-school students competing for a place at university, the bizarre scenario came true in March. Newspaper accounts of the cipher, and rumours that a privileged few had been told how to decode it, sparked student demonstrations across the country. Although the cipher may have been an accident of the test’s design rather than a conspiracy, outrage has persisted and the debacle has become an issue in the run-up to the country’s general election on 12 June.

Last week, prime minister Recep Tayyip Erdoğan accused a journalist investigating the case of campaigning to discredit the government. “He will pay a heavy price, not now but in the future,” said Erdoğan. This is no empty threat — several investigative journalists were arrested in March as part of an inquiry into an alleged plot to overthrow the government.

Ankara’s Chief Public Prosecutor’s Office declared on 11 May that the cipher indisputably exists. It said it had found no evidence of cheating, but requested permission to investigate Ali Demir, head of the Student Selection and Placement Center (ÖSYM), the Council of Higher Education office responsible for the examinations. On 25 May, the council’s president Yusuf Ziya Özcan said that it would probably not consider the investigation request until after the election.

In a country where corruption is rife, some educators see benefit in multiple-choice examinations because they can be graded by computer. In previous years, up to ten different versions of the exam paper had been printed with the same questions in a different order, to reduce the possibility of copying. Crucially, the easiest questions had

always appeared first. This year, however, ÖSYM decided to create a unique version for each student by changing the orders of both questions and answers, commissioning an external company to create an algorithm for the task.

According to some academic experts, that algorithm may have inadvertently generated the cipher that eventually emerged for those questions with numerical answers. If the answers were rearranged in ascending order and placed next to the original list, the correct answer occupied the same position in each list. Enough questions followed the pattern that students could have passed the

exam by relying entirely on this decryption method.

“It’s not possible to tell” whether the cipher was accidental or deliberate, says Ersan Akyıldız, a cryptologist at the Middle East Technical University in Ankara. Either way, he is critical of the test design, noting that the reshuffling of questions put students who encountered more daunting questions at the beginning of the exam at

a disadvantage. The ÖSYM team should have consulted academic experts in education and cryptography before trying such an ambitious experiment, Akyıldız says.

Many Turkish academics also complain that multiple-choice exams do not test reasoning skills well enough to be suitable for selecting university students. “We need to have real university entrance exams in Turkey, like the [international] baccalaureate,” says chemical engineer Kemal Gürüz, a former president of the Higher Education Council, at Middle East Technical University.

The stakes are high. Turkey has a policy of widening access to higher education but has only 600,000 university places available this year. As a consequence, profitable cram schools have flourished — and have lobbied against change. ■



Turkish premier Recep Tayyip Erdoğan threatened a journalist over an exam row.

REUTERS/M. AZAKIR

➔ **NATURE.COM**
For more on the region, read *Nature Middle East*.
nature.com/middleeast

First sale for quantum computing

But critics say that D-Wave's system is still something of a black box.

BY ZEEYA MERALI

It could turn out to be a milestone for quantum computing. Last week, D-Wave Systems of Burnaby in British Columbia, Canada, announced the first sale of a commercial quantum computer, to global security firm Lockheed Martin, based in Bethesda, Maryland.

Yet perhaps fittingly for a quantum device, uncertainty persists around how the impressive black monolith known as D-Wave One actually works. Computer scientists have long questioned whether D-Wave's systems truly exploit quantum physics, and although the company last month published a paper in *Nature* (M. W. Johnson *et al. Nature* **473**, 194–198; 2011) to help verify its quantum credentials, some say that the technique used is still in doubt.

Quantum computers could revolutionize the way we tackle problems that stump even the best classical computers, which store and process their data as 'bits' — essentially a series of switches that can be either on or off.

The power of quantum bits — or qubits — is that they can be on and off simultaneously. Connect enough qubits together using quantum entanglement and a computer should be able to zip through a multitude of calculations in parallel, at astonishing speed.

But quantum computers are notoriously difficult to construct, with most research groups struggling to entangle more than a handful of qubits. So D-Wave's claims to have a working 128-qubit processor — combined with the company's former reluctance to publish details of the technique in peer-reviewed journals — have long raised eyebrows, explains Scott Aaronson, a computer scientist at the Massachusetts Institute of Technology in Cambridge.

This did not deter Lockheed Martin, however. Communications manager Thad Madden says that the company spent a year

reviewing the D-Wave One computer before purchasing it. Although D-Wave has been cagey about specific applications, Lockheed Martin plans to use the technology to help it build "cyber-physical systems", which integrate software with environmental sensors, says Madden.

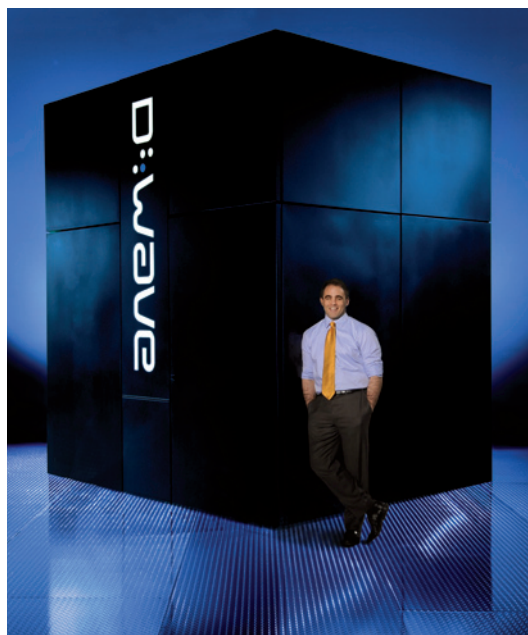
D-Wave's co-founder, Geordie Rose, says that the sale demonstrates that quantum computing is finally living up to its decades-long promise. Aaronson, however, thinks that the computer-science community will need more

driven by plain old classical physics. At the time, D-Wave did not respond with any publications ruling out this possibility.

But the company's paper in *Nature* demonstrated definitive quantum behaviour in a system with eight qubits, made from superconducting niobium loops. Magnetic fields manipulate the combined energy state of the qubits until the system defines the parameters of the problem to be solved. Because the system exists in a quantum superposition, it can simultaneously 'search' through multiple energy states that each represent different solutions, explains Rose. Cooling the system snaps it out of the superposition, so that it settles into a single, low-energy state that represents the final answer, an approach known as quantum annealing (W. D. Oliver *Nature* **473**, 164–165; 2011). D-Wave One uses sixteen of these eight-qubit cells.

Sceptics had argued that the system could actually work by using thermal fluctuations — rather than quantum effects — to jostle the qubits through various energy states. However, because these classical effects are significant only above 45 millikelvin, and D-Wave's latest paper shows that its system works below that barrier, only quantum effects can be responsible.

Aaronson says that the paper is a step in the right direction, but only a small one. "There is an enormous gap between demonstrating some kind of quantum effect in eight qubits, as they have done here, and saying that they have a 128-qubit chip that can perform a computationally interesting task faster than a conventional computer," he says. He notes that D-Wave has not demonstrated that the qubits are entangled, a property that most physicists think is essential for quantum computing. "I genuinely hope that D-Wave succeeds," he adds. "But I will only believe it has when I see evidence for it." ■ **SEE FEATURE P.24**



D-Wave co-founder Geordie Rose with his black monolith.

convincing. "Just because a flagship company has bought the system, doesn't mean that it now works," he says.

That mistrust goes back to 2007, when D-Wave apparently demonstrated a 16-qubit computer that could solve a Sudoku puzzle. Many computer scientists and physicists suggested that the device was actually being

DOMINIC SCHAEFER PHOTOGRAPHY

TOP STORY

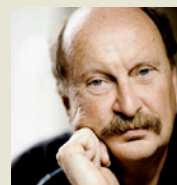


United States field-tests terrorist 'pre-crime' detector
go.nature.com/d4uwyy

OTHER NEWS

- Smart-REDD plan targets causes of deforestation go.nature.com/udpmwl
- Link between biomarkers and disease is often exaggerated go.nature.com/vog9zg
- Roman ship had on-board fish tank go.nature.com/bhhgyk

Q&A



Edzard Ernst, the world's first professor of alternative medicine, is stepping down
go.nature.com/ojxiyo



Felisa Wolfe-Simon has been criticized for her work on bacteria recovered from an arsenic-rich lake.

MICROBIOLOGY

Will you take the 'arsenic-life' test?

Critiques prompt researchers to offer samples of poison-tolerant microbe to doubters.

BY ERIKA CHECK HAYDEN

At first, it sounded like the discovery of the century: a bacterium that can survive by using the toxic element arsenic instead of phosphorus in its DNA and in other biomolecules.

But scientists have lined up to criticize the claim since it appeared in *Science* six months ago¹. Last week, the journal published a volley of eight technical comments^{2–9} summarizing the key objections to the original paper, along with a response from the authors¹⁰, who stand by their work.

The authors of the original paper are also offering to distribute samples of the bacterium, GFAJ-1, so that others can attempt to replicate their work. The big question is whether researchers will grab the opportunity to test such an eye-popping claim or, as some are already saying, they will reject as a waste of time the chance to repeat work they believe is fundamentally flawed. "I have not found anybody outside of that laboratory who supports the work," says Barry Rosen of Florida International University in Miami, who published an earlier critique of the paper¹¹.

Some are also frustrated that the authors did not release any new data in their response, despite having had ample time to conduct follow-up experiments of their own to bolster their case. "I'm tired of rehashing these preliminary data," says John Helmann of Cornell University in Ithaca, New York, who critiqued the work in January on the Faculty of 1000 website¹². "I look forward to the time when they or others in the field start doing the sort of rigorous experiments that need to be done to test this hypothesis."

The original study¹, led by Felisa Wolfe-Simon, a NASA astrobiology research fellow at the US Geological Survey in Menlo Park, California, looked at bacteria taken from the arsenic-rich Mono Lake in southern California. The authors grew the bacteria in their lab using a medium that contained arsenic but no phosphorus. Even without this essential element of life, the bacteria reproduced and integrated arsenic into their DNA to replace the missing phosphorus, the paper reported.

➔ NATURE.COM
For a longer version
of this story see:
go.nature.com/oaetjq

"We maintain that our interpretation of As [arsenic] substitution,

based on multiple congruent lines of evidence, is viable," Wolfe-Simon and her colleagues wrote in last week's response¹⁰.

But critics have pointed out that the growth medium contained trace amounts of phosphorus^{2,3} — enough to support a few rounds of bacterial growth⁵. They also note that the culturing process could have helped arsenic-tolerant bacteria to survive by killing off less well-equipped microbes³.

Others say that there is simply not enough evidence that arsenic atoms were incorporated into the bacterium's DNA^{4,6–9}. The chemical instability of arsenate relative to phosphate makes this an extraordinary claim that would "set aside nearly a century of chemical data concerning arsenate and phosphate molecules", writes Steven Benner⁴ of the Foundation for Applied Molecular Evolution in Gainesville, Florida.

A leading critic of the work, Rosemary Redfield of the University of British Columbia in Vancouver, Canada, says that it would be "relatively straightforward" to grow the bacteria in arsenic-containing media and then analyse them using mass spectrometry to test whether arsenic is covalently bonded into their DNA backbone.

Redfield says that she will probably get samples of GFAJ-1 to run these follow-up tests, and hopes that a handful of other laboratories will collaborate to repeat the experiments independently and publish their results together.

But some principal investigators are reluctant to spend their resources, and their students' time, replicating the work. "If you extended the results to show there is no detectable arsenic, where could you publish that?" asks Simon Silver of the University of Illinois at Chicago. "How could the young person who was asked to do that work ever get a job?"

Helmann says that he is in the process of installing a highly sensitive mass spectrometer that can measure trace quantities of elements, which could help refute or corroborate the findings. But the equipment would be better employed on original research, he says. "I've got my own science to do." ■

1. Wolfe-Simon, F. *et al. Science* doi:10.1126/science.1197258 (2010).
2. Foster, P. L. *Science* doi:10.1126/science.1201551 (2011).
3. Redfield, R. J. *Science* doi:10.1126/science.1201482 (2011).
4. Benner, S. A. *Science* doi:10.1126/science.1201304 (2011).
5. Cotner, J. B. & Hall, E. K. *Science* doi:10.1126/science.1201943 (2011).
6. Schoepp-Cothenet, B. *et al. Science* doi:10.1126/science.1201438 (2011).
7. Csabai, I. & Szathmáry, E. *Science* doi:10.1126/science.1201399 (2011).
8. Oehler, S. *Science* doi:10.1126/science.1201381 (2011).
9. Borhani, D. W. *Science* doi:10.1126/science.1201255 (2011).
10. Wolfe-Simon, F. *et al. Science* doi:10.1126/science.1202098 (2011).
11. Rosen, B. P., Ajees, A. A. & McDermott, T. R. *BioEssays* **33**, 350–357 (2011).
12. Helmann, J. *F1000.com*/6854956 (2011).

A s best he can remember, says Paul Davies, the telephone call that changed his professional life came some time in November 2007, as he was sitting in the small suite of offices that comprise his Beyond Center at Arizona State University (ASU) in Tempe.

Until then, the questions that animated Davies' research and 19 popular-science books had grown out of his training in physics and cosmology: how did the Universe come to exist? Why are the laws of physics suited for life? What is time? And how did life begin? But this particular call was nothing to do with that. The caller — Anna Barker, then the deputy director of the US National Cancer Institute (NCI) in Bethesda, Maryland — explained that she needed his help in the 'War on cancer'. Forty years into the government's multibillion-dollar fight, said Barker, cancer survival rates had barely budged. The hope now was that physicists could bring some radical new ideas to the table, and she wanted Davies to give a keynote address at an NCI workshop explaining how.

Ummm, sure, said Davies, who until that minute had been only vaguely aware that the NCI existed. "But I don't know anything about cancer."

"That's okay," Barker replied. "We're after fresh insights."

And with that, says Davies, he was hooked. "If it had been just, 'Give us another beam, I wouldn't have been interested,' he says, referring to X-rays, particle beams, magnetic resonance imaging and the many other tools that physicists had provided to medicine. But an opportunity to contribute entirely new concepts and ways of thinking — "now that", says Davies, "was exciting".

That excitement explains how the 65-year-old Davies, at an age when most academics are planning their retirement, finds himself embarking on practically a new career. Barker's original workshop metamorphosed into a network of 12 Physical Sciences-Oncology Centers, which launched in late 2009. Davies now finds himself the principal investigator of one such centre, and a major player in the physics-meets-cancer effort as a whole. Cancer gives Davies a new realm in which to exercise what many colleagues regard as his greatest talent: asking 'dumb' questions that provoke fresh ways of thinking about a problem. "Paul is wrong sometimes. But he is not afraid to ask a very naive question that gets at the heart of the matter," says Robert Austin, a biophysicist who heads another of the 12 centres, at Princeton University in New Jersey. Davies' questions have addressed topics ranging from metastasis (when tumour cells come apart and migrate, is it because of some physical change in their stickiness?) to subatomic physics (is cancer influenced by quantum effects inside biomolecules?). "I often joke that my main qualification for cancer research is that I am unencumbered by any prior knowledge of the subject," Davies says.

True, his naivety sometimes makes biologists grit their teeth. ("Aaargh! Physicists!" wrote Paul 'PZ' Myers, a biologist at the University of Minnesota, Morris, in a blog response to Davies' proposal earlier this year that tumours are a reversion to primitive genetic mechanisms that pre-date the dawn of multicellular life.) "But his critics don't appreciate the value of a disruptive agent," says biophysicist Stuart Lindsay, who works closely with Davies at the ASU physics-cancer centre. "It takes someone like Paul, constantly nagging, asking disruptive questions, to get people to take a fresh look at their assumptions."

Davies says that he has been asking questions as long as he can remember. The suburbs of London were a dull place to grow up, he explains, thinking back to the post-Second-World-War austerity that prevailed in the years after he was born there in 1946. "No toys. No money. We made our own entertainment — so

we had to use our imagination a lot."

Maybe that's why he became so fascinated with shooting stars and astronomy, he says. "I liked the fact that, just by looking up, you could escape into this wonderland out there." And maybe that's why by the

age of ten he had become enthralled with atoms, which seemed to embody a hidden order behind the surface complexity of the Universe. A few years later, he says, "I remember being struck by the fact that the brain is made of atoms, and atoms follow the laws of physics — so how can we have free will?" By age 16, says Davies, his course was set: he would become a theoretical physicist and spend his life trying to answer the "deep questions".

He began to explore one such question — how does quantum theory operate when space and time are curved by gravity? — first as a PhD student at University College London in the late 1960s and later as a lecturer at Kings College London. "It was a connection between the very small and the very large, between quantum mechanics and the whole Universe," Davies explains. He eventually summarized the field's accomplishments as co-author of a classic monograph, *Quantum Fields in Curved Space* (1982). But even then his interests were not easily confined — and he was beginning to develop a parallel outlet for them. In the early 1970s, the British magazine *Physics Bulletin* invited Davies to write a popular article dealing with a long-standing conundrum he had touched on in his PhD dissertation: why does time seem to flow only in one direction — towards the future — even though most physical laws make no distinction? After the article appeared, a publisher asked him for a book on the subject. And yet "I barely scraped through English class!" Davies says.

QUESTIONS IN WRITING

But Davies discovered that he enjoyed popular writing, and had a knack for it. Whereas early titles mirrored Davies' research in pure physics and cosmology, dealing with the physics of black holes, the unification of forces, and quantum theory, later ones have reflected his widening interests. In 1980, he joined the physics department at the University of Newcastle, UK, where half the department was focused on geophysics. So just by osmosis, he began to learn about the long history of life on our planet, and he found himself inexorably drawn to another big question: how is life even possible? How could such complexity arise in a lifeless Universe, purely by the action of natural law? Davies explored the complexity question in popular books such as *The Cosmic Blueprint* (1987), and then — after moving to the University of Adelaide, Australia, in 1990 — went on to tackle the philosophical implications of extraterrestrial life in books such as *Are We Alone?* (1995). At the same time, titles such as *God and the New Physics* (1983) and *The Mind of God* (1992) revealed his willingness to engage in the dialogue about religion and science — efforts that he assumes contributed to his winning the 1995 John Templeton Prize for Progress in Religion and, Davies admits, earned him plenty of criticism from anti-religious scientists.

But Davies has never let critics stop him from asking provocative questions. In the 1990s, for example, he began to wonder whether Earth and Mars might share a biosphere. Might chunks of rock blasted loose from one planet by ancient asteroid impacts have carried viable microbes to the other when the rocks fell as meteorites? "Most people dismissed the idea as total nonsense, in rather blunt terms," says Davies. By 2004, he was wondering whether life could have originated on our planet more than once, with each lineage based on utterly different biochemistries, perhaps even without DNA or RNA. Might some of those alien lineages still be alive today? That question has led a number

The disruptor

BY M. MITCHELL WALDROP

Paul Davies likes to ask big questions.

But how did the free-thinking cosmologist suddenly find himself probing the physics of cancer?



of astrobiologists to search for alternative life forms in harsh environments such as Mono Lake in California. It is also how Davies came to act as an adviser and co-author on last year's highly controversial paper claiming to have found bacteria that break life's rules by using arsenic instead of phosphorus in their DNA (F. Wolfe-Simon *et al.* *Science* doi:10.1126/science.1197258; 2010).

In 2004, Davies was contacted by Michael Crow, president of ASU, who was looking to overhaul the university's department-based hierarchy in favour of a more interdisciplinary approach (see *Nature* **446**, 968–970; 2007). Davies seemed to be one of those rare thinkers who could shake up the status-quo thinking in academia, says Crow. "Individuals who are polymaths, able to think across different subjects, and who personify 'disruptive thinking'." He contacted Davies at Macquarie University in Sydney, where Davies had helped to found the Australian Centre for Astrobiology in 2001, and offered to set him up in a centre where he could freely pursue all his interests.

It was an offer too good to refuse, says Davies. He moved to ASU to head the Beyond Center for Fundamental Concepts in Science in September 2006. Just over a year later, Anna Barker was on the other end of the phone.

MAKING A LIST

Once Davies had signed up to talk at the physics-cancer workshop, he had to figure out what to say. He knew he couldn't tell experienced cancer researchers how to do their jobs. So instead, he did what he does best: "I made a list of dumb questions."

First was whether physics could contribute anything at all to cancer research. A little reading and talking to colleagues convinced him it could. "To my astonishment, I learned that physical forces can affect gene expression," he says. Stretching, squashing — lots of things would do the trick.

Davies also learned that biologists rarely think about the cell as a physical object. "Look at something as straightforward as 'where does metastasis occur?'" says Davies. Does a tumour tend to seed itself in a second organ simply because blood flows there from the primary site? And what makes tumour cells suddenly break apart and become mobile in the first place, despite all the biophysical forces that tend to stick them together? He added those to the list.

Next, Davies was struck by the fact that biologists can now explore the cell in enormous detail — practically molecule by molecule. But that very power, he says, has often beguiled cancer researchers into focusing on individual genes and all the other pieces that go wrong, instead of how the pieces come together into a complex whole. "It's like trying to run the economy of the United States by measuring every transaction in every commodity and every city," says Davies. Granted, the comparatively new discipline of systems biology has been trying to take a more global view. But even so, says Davies, few cancer biologists are familiar with non-linear systems analysis, network theory or any of the other tools that have been developed by mathematicians and physicists over the past few decades to deal with complex systems.

The questions, he says, were "all very, very basic. My level of ignorance was embarrassing."

But when the workshop convened in Arlington, Virginia, on 26 February 2008, Davies' talk was a hit. Barker remembers being delighted. "He gave a fascinating perspective, at a level biologists often just haven't thought about." And the cancer researchers in the audience were very receptive, recalls

Austin, who also presented at the workshop. Even though many biologists object to the involvement of physicists in their field, "people in the oncology community have been very welcoming," he says. "They know they have a problem."

In December 2008, the NCI outlined its plan to fund the 12 physics-oncology centres at roughly US\$2 million apiece over five years, and invited applications to host them. Each one would look at cancer from one of four points of view: physics, evolution, biological-information processing or complex systems.

Researchers at ASU's four-year-old Biodesign Institute spearheaded an application. William Grady, a gastroenterologist with a joint appointment at the University of Washington Medical Center and at the Fred Hutchinson Cancer Center in Seattle, agreed to serve as the senior scientific investigator on the project. But by the NCI's rules, the proposal needed a principal investigator in the physical sciences.

Davies describes himself as a reluctant draftee to that role. "I'm not a natural administrator," he says. But every time he protested "I have no credibility", his colleagues would point to his reception at the workshops and insist "You've got lots." When the NCI announced its selection of the 12 centres in October 2009, ASU was among them.

Since then, Davies has tailored his contribution to his strengths: he runs about three workshops a year for participants throughout the physics-oncology community. The goal is to trigger new collaborations, new experiments, new thinking — and the topics have ranged "all the way from downright crazy to productive", says Lindsay. Quantum effects in cancer might fall in the former category; the latter might include the physics of chromatin, the mass of DNA and protein in the cell nucleus.

"He won't allow people to get into a shoot-out of your theory versus mine. Just lots of critical thinking: 'what does this mean?'," says a frequent participant, oncologist Donald Coffey of the Johns Hopkins University in Baltimore, Maryland. "These are dynamite meetings. I come out incredibly enlightened, with lots of things to think about. And it works because of his personality."

DYNAMITE MEETINGS

What remains to be seen, of course, is whether the multimillion-dollar physics-cancer effort will pay off. The centres are only now starting to produce their first papers, Lindsay says, "and there's nothing earth-shattering yet". Several of the papers, including various studies on the mechanical properties of cancer cells and chromosomes along with a formal write-up from Davies on the idea that irritated Myers, appear in the February 2011 issue of the journal *Physical Biology*. It's early days, agrees Austin — but not too early to fret. "I worry that there aren't enough Paul Davies around to ask disruptive things," he says. "I worry that we'll become conventional —

another failed assault on cancer. I really hate the thought that in ten years we'll find we haven't accomplished anything."

Davies isn't worried — and he is happy to keep asking those disruptive questions. "My mother didn't understand science at all," he says, by way of an explanation. In fact, he was the first person in his family to go to university. "But she was fond of saying that she hoped I could do two things with it." One was to build a robot to help with housework — something he'll have to leave to others. But the other was to find a cure for cancer.

"She would be quite thrilled that I'm finally doing something useful." ■

M. Mitchell Waldrop is a features editor for *Nature*.

"I worry that there aren't enough Paul Davies around to ask disruptive things."



THE POWER OF DISCORD

BY ZEEYA MERALI

Physicists have always thought quantum computing is hard because quantum states are incredibly fragile. But could noise and messiness actually help things along?

In 2008, quantum physicist Andrew White found himself building a “ridiculous machine” in his lab at the University of Queensland in Brisbane, Australia.

White had spent years working on quantum computation, attempting to exploit subatomic physics to create a device with the potential to outperform its best macroscopic counterparts. And he had learned that it was a tough job: the required quantum systems are fragile, and demand immaculate laboratory conditions to survive long enough to be of any use. Now White was setting out to test an unorthodox quantum algorithm that seemed to turn that lesson on its head. In this scheme, messiness and disorder would be virtues, not vices — and perturbations in the quantum system would drive computation, not disrupt it.

“I honestly thought, there’s no way this will work,” says White. But when he turned his ridiculous machine on, it ran¹.

White’s experiment is just one of several in recent years that have suggested a fresh approach to quantum computers. The conventional view is that such devices should get their computational power from quantum entanglement — a phenomenon through which particles can share information even when they are separated by arbitrarily large distances. But the latest experiments suggest that entanglement might not be needed after all. Algorithms could

instead tap into a quantum resource called discord, which would be far cheaper and easier to maintain in the lab. More experiments will be required to convince the many sceptics that the approach will work. But if it pans out, the era of widespread quantum computation could arrive sooner than anyone expected.

UNPRECEDENTED SPEED

The idea of quantum computing dates from the 1980s, when Nobel-prizewinning physicist Richard Feynman realized that a machine using quantum rules could whizz through calculations that would take a standard computer billions of years. Classical computers have to encode their data in an either/or fashion: each bit of information takes a value of 0 or 1, and nothing else. But the quantum world is the realm of both/and. Particles can exist in ‘superpositions’ — occupying many locations at the same time, say, or simultaneously spinning clockwise and anticlockwise.

So, Feynman argued, computing in that realm could use quantum bits of information — qubits — that exist as superpositions of 0 and 1 simultaneously. A string of 10 such qubits could represent all 1,024 10-bit numbers simultaneously. And if all the qubits shared information through entanglement, they could race through myriad calculations in parallel — calculations that their classical counterparts would have to plod through sequentially (see ‘Quantum computing’).

M. WEINBERGER

➔ **NATURE.COM**
For more on possible routes to quantum computing, visit:
go.nature.com/fdxdio

The notion that quantum computing can be done only through entanglement was cemented in 1994, when Peter Shor, a mathematician at the Massachusetts Institute of Technology in Cambridge, devised an entanglement-based algorithm² that could factorize large numbers at lightning speed — potentially requiring only seconds to break the encryption currently used to send secure online communications, instead of the years required by ordinary computers. In 1996, Lov Grover at Bell Labs in Murray Hill, New Jersey, proposed an entanglement-based algorithm³ that could search rapidly through an unsorted database; a classical algorithm, by contrast, would have to laboriously search the items one by one.

But entanglement has been the bane of many a quantum experimenter's life, because the slightest interaction of the entangled particles with the outside world — even with a stray low-energy photon emitted by the warm walls of the laboratory — can destroy it. Experiments with entanglement demand ultra-low temperatures and careful handling. “Entanglement is hard to prepare, hard to maintain and hard to manipulate,” says Xiaosong Ma, a physicist at the Institute for Quantum Optics and Quantum Information in Vienna. “It has been thoroughly investigated for years, with people expending much time and effort, but achieving little efficiency.” The current entanglement record-holder intertwines just 14 qubits (ref. 4), yet a large-scale quantum computer would need several thousand. Any scheme that bypasses entanglement would be warmly welcomed, says Ma.

Clues that entanglement isn't essential after all began to trickle in about a decade ago, with the first examples of rudimentary quantum computation. In 2001, for instance, physicists at IBM's Almaden Research Center in San Jose and Stanford University, both in California, used a 7-qubit system to implement Shor's algorithm⁵, factorizing the number 15 into 5 and 3. But controversy erupted over whether the experiments deserved to be called quantum computing, says Carlton Caves, a quantum physicist at the University of New Mexico (UNM) in Albuquerque.

The trouble was that the computations were done at room temperature, using liquid-based nuclear magnetic resonance (NMR) systems, in which information is encoded in atomic nuclei using an internal quantum property known as spin. Caves and his colleagues had already shown⁶ that entanglement could not be sustained in these conditions. “The nuclear spins would just be jostled about too much for them to stay lined up neatly,” says Caves. According to the orthodoxy, no entanglement meant no quantum computation.

The NMR community gradually accepted that they had no entanglement, says Jiangfeng Du, an NMR-computing specialist at the University of Science and Technology of China, in Hefei. Yet the computations were producing real results. In 2001, Du and his colleagues

published the first experiment to explicitly perform a quantum search without exploiting entanglement⁷.

“These experiments really called into question what gives quantum computing its power,” says Animesh Datta, a physicist at the University of Oxford, UK. If researchers hope to build a large-scale quantum computer, they need to understand how the computation works.

ORDER OUT OF DISCORD

Datta, at the time a graduate student supervised by Caves at UNM, began to search for an alternative explanation. He came across discord, an obscure measure of quantum correlations first proposed⁸ in 2000 by Wojciech Zurek, a quantum physicist at the Los Alamos National Laboratory in New Mexico. Discord quantifies how much a system can be disrupted when people observe it to gather information. Macroscopic systems are not affected by observation, and so have zero discord. But quantum systems are unavoidably affected because measurement forces them to settle on one of their many superposition values, so any possible quantum correlations, including entanglement, give a positive value for discord.

The concept was largely ignored for years because it seemed so abstract, says Vlatko Vedral, a quantum physicist at the University of Oxford, who in 2002 independently derived a

mixed qubit will yield a random result, providing little or no useful information.

But Knill and Laflamme pondered what would happen if a mixed qubit was sent through an entangling gate with a pure qubit. The two could not become entangled but, the physicists argued, their interaction might be enough to carry out a quantum computation, with the result read from the pure qubit. If it worked, experimenters could get away with using just one tightly controlled qubit, and letting the others be battered by environmental noise and disorder. “It was not at all clear why that should work,” says White. “It sounded as strange as saying they wanted to measure someone's speed by measuring the distance run with a perfectly metered ruler and measuring the time with a stopwatch that spits out a random answer.”

Datta supplied an explanation¹¹. With Caves and Anil Shaji, a physicist then at UNM, he calculated that the computation could be driven by the quantum correlation between the pure and mixed qubits — a correlation given mathematical expression by the discord.

It was a bold claim, says Kavan Modi, an expert on discord at the Centre for Quantum Technologies at the National University of Singapore. “Before that, if you announced that discord was as important for computation as entanglement — if not more so — at a confer-

“Discord could be like sunlight, which is plentiful but has to be harnessed in a certain way to be useful.”

mathematical expression for discord⁹ in collaboration with Leah Henderson at the University of Bristol, UK. “But that changed when Datta connected discord to quantum computing.”

Datta had seized on an algorithm¹⁰ proposed a few years earlier by NMR researchers Emanuel Knill, now at the US National Institute of Standards and Technology in Boulder, Colorado, and Raymond Laflamme, now at the University of Waterloo in Canada. Knill and Laflamme challenged the idea that quantum computing requires physicists to painstakingly prepare a set of pristine qubits in the lab.

In a typical optical experiment, the pure qubits might consist of horizontally polarized photons representing 1 and vertically polarized photons representing 0. Physicists can entangle a stream of such pure qubits by passing them through a processing gate such as a crystal that alters the polarization of the light, then read off the state of the qubits as they exit. In the real world, unfortunately, qubits rarely stay pure. They are far more likely to become messy, or ‘mixed’ — the equivalent of unpolarized photons. The conventional wisdom is that mixed qubits are useless for computation because they cannot be entangled, and any measurement of a

ence, people would laugh out loud at you.” But it seemed shocking only because, at the time, physicists had never really analysed computation in real-world scenarios that included mixed states. “It's true that you must have entanglement to compute with idealized pure qubits,” says Modi. “But when you include mixed states, the calculations look very different.”

Datta and his colleagues presented experimenters with a testable discord-based scheme. White doubted it would work, but jumped at the prospect of trying it out. “I'm a lazy experimenter, so I loved the thought of quantum computation without the hassle of entanglement,” he laughs.

White was already practised at using polarized photons. He ran the computation as prescribed by Datta and, by averaging the values of the pure qubit over 2,000 runs, successfully summed the diagonal elements of a 2×2 matrix of numbers¹. “It's a small matrix, but this was a proof-of-principle to show that you get the right answer in a reasonable number of runs, as predicted,” says White.

The team confirmed that the qubits were not entangled at any point. Intriguingly, when the researchers tuned down the polarization

quality of the one pure qubit, making it almost mixed, the computation still worked. “Even when you have a system with just a tiny fraction of purity, that is vanishingly close to classical, it still has power,” says White. “That just blew our minds.” The computational power only disappeared when the amount of discord in the system reached zero. “It’s counter-intuitive, but it seems that putting noise and disorder in your system gives you power,” says White. “Plus, it’s easier to achieve.”

For Ma, White’s results provided the “wow! moment” that made him take discord seriously. He was keen to test discord-based algorithms that used more than the two qubits used by White, and that could perform more glamorous tasks, but he had none to test. “Before

I can carry out any experiments, I need the recipe of what to prepare from theoreticians,” he explains, and those instructions were not forthcoming.

Although it is easier for experimenters to handle noisy real-world systems than pristine ones, it is a lot harder for theoretical physicists to analyse them mathematically. “We’re talking about messy physical systems, and the equations are even messier,” says Modi. For the past few years, theoretical physicists interested in discord have been trying to formulate prescriptions for new tests. Such experiments are essential if advocates of discord are to win over the wider physics community, says Antonio Acín, a quantum physicist at the Institute of Photonic Sciences in Barcelona, Spain. He

notes that no one has yet proved that discord is essential to computation — just that it is there. Rather than being the engine behind computational power, it could just be along for the ride, he argues. Last year, Acín and his colleagues calculated that almost every quantum system contains discord¹². “It’s basically everywhere,” he says. “That makes it difficult to explain why it causes power in specific situations and not others.”

Modi shares the concern. “Discord could be like sunlight, which is plentiful but has to be harnessed in a certain way to be useful. We need to identify what that way is,” he says.

Du and Ma are independently conducting experiments to address these points. Both are attempting to measure the amount of discord at each stage of a computation — Du using liquid NMR and electron-spin resonance systems, and Ma using photons. They hope to have results by the end of the year.

A finding that quantifies how and where discord acts would strengthen the case for its importance, says Acín. But if these tests find discord wanting, the mystery of how entanglement-free computation works will be reopened. “The search would have to begin for yet another quantum property,” he adds.

Vedral notes that even if Du and Ma’s latest experiments are a success, the real game-changer will be discord-based algorithms for factorization and search tasks, similar to the functions devised by Shor and Grover that originally ignited the field of quantum computing. “My gut feeling is that tasks such as these will ultimately need entanglement,” says Vedral. “Though as yet there is no proof that they can’t be done with discord alone.”

Zurek says that discord can be thought of as a complement to entanglement, rather than as a usurper. “There is no longer a question that discord works,” he declares. “The important thing now is to find out when discord without entanglement can be exploited most usefully, and when entanglement is essential.” ■ [SEE NEWS P.18](#)

QUANTUM COMPUTING

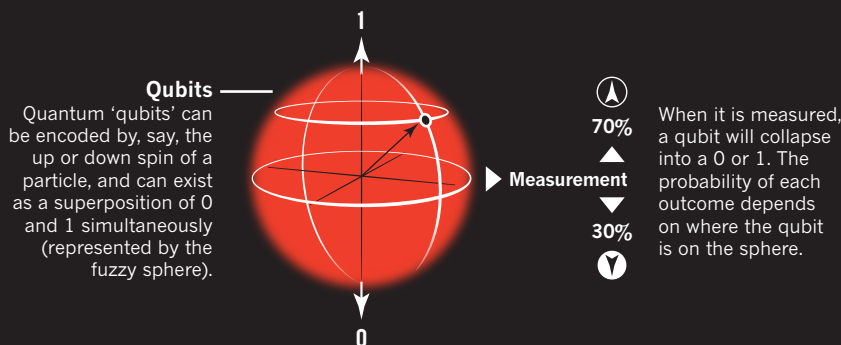
Devices based on subatomic physics could make calculations far faster than conventional machines — if nothing spoils their quantum weirdness.

1. SUPERPOSITION



Bits

A classical computer encodes information in strings of ‘bits’, which can take one of two values: 0 or 1.



2. QUANTUM COMPUTATION USING ENTANGLEMENT

Before computation

Data are spread across entangled qubits, which are isolated from the environment.

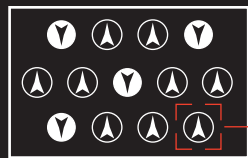


Perform computation



After computation

The entangled qubits have processed their information in parallel.



One qubit serves as a spokesman. Taking an average of measurements (0 or 1) over many runs gives the answer.

3. QUANTUM COMPUTATION USING DISCORD

Before computation

Only one qubit is protected from the environment.

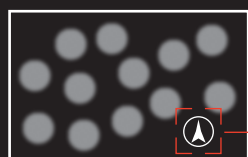


Perform computation



After computation

The other qubits have been exposed to noise and disruption.



Surprisingly, measuring the protected qubit and averaging over many runs still gives the right answer.

Zeeva Merali is a freelance writer based in London.

1. Lanyon, B. P., Barbieri, M., Almeida, M. P. & White, A. G. *Phys. Rev. Lett.* **101**, 200501 (2008).
2. Shor, P. in *Proc. 35th Annual Symp. Foundations Comp. Sci.* 124–134 (IEEE Press, 1994).
3. Grover, L. K. in *Proc. Twenty-Eighth Annual ACM Symp. Theory Comput.* 212–219 (ACM, 1996).
4. Monz, T. et al. *Phys. Rev. Lett.* **106**, 130506 (2011).
5. Vandersypen, L. M. K. et al. *Nature* **414**, 883–887 (2001).
6. Braunstein, S. L., Caves, C. M., Josza, R., Linden, N., Popescu, S. & Shack, R. *Phys. Rev. Lett.* **83**, 1054–1057 (1999).
7. Du, J. et al. *Phys. Rev. A* **64**, 042306 (2001).
8. Zurek, W. H. *Ann. der Physik (Leipzig)* **9**, 853–862 (2000).
9. Henderson, L. & Vedral, V. *J. Phys. A* **34**, 6899 (2002).
10. Knill, E. & LaFlamme, R. *Phys. Rev. Lett.* **81**, 5672–5675 (1998).
11. Datta, A., Shaji, A. & Caves, C. M. *Phys. Rev. Lett.* **100**, 050502 (2008).
12. Ferraro, A., Aolita, L., Cavalcanti, D., Cucchietti, F. M. & Acín, A. *Phys. Rev. A* **81**, 052318 (2010).

COMMENT

BEHAVIOUR Thirty years since first AIDS cases, denial still hampers the fight **p.29**

EVOLUTION The power of trial and error in economic and social progress **p.32**

BRAIN Memory supplies cognition's building blocks **p.34**

OBITUARY Physicist who changed imaging, Willard Boyle, remembered **p.37**



G. PIROZZI/PANOS



These orphans in Zimbabwe have looked after themselves since their parents died of AIDS.

Children of the AIDS pandemic

Practical support and psychosocial interventions are desperately needed to help those dealing with the fallout of AIDS, says **Lucie Cluver**.

I'm sitting on a piece of corrugated iron in a dusty South African township, with the smell of burning rubber in the air. A set of graphs comes to life for me. A 15-year-old girl traces patterns in the sand as she tells me how she found out she is HIV-positive. Lindiwe is one of 12 million children in sub-Saharan Africa orphaned as a result of AIDS. Her parents were said to have died of 'TB and bewitchment', but their symptoms confirmed that their deaths were among the 850 caused each day by AIDS in South Africa.

HIV/AIDS turns 30 this year, and so will a million or so of the children it has orphaned. It is not yet clear what the epidemic's long-term impact on this generation has been, or what it will be for all the other youths affected. Lindiwe is part of the developing world's first longitudinal study of the impact on children of parents made ill or killed by AIDS.

Beginning in 2005, a group of social scientists, psychologists and social workers followed 1,000 girls and boys in South Africa over four years — including those orphaned

by AIDS, those orphaned by other causes and those whose parents were still alive. As study director, I had never anticipated that our team's repeated visits would make us confidantes for these children, whose opportunities to talk about the disease are constrained by stigma. I had also underestimated the extent to which their lives would offer a window onto the complexities of the HIV epidemic.

LONG SUFFERING

For 30 years, the scientific community has fought the war against HIV/AIDS on many fronts: prevention, treatment, the elusive cure. Much has been achieved: 37% of sub-Saharan Africans who need antiretrovirals are receiving them compared with 2% in 2003.

But research is now revealing a new and daunting battlefield — the multiple social consequences of AIDS on families, especially children¹. With more than 22 million people in sub-Saharan Africa infected with HIV, many of whom are parents, some 70 million children are likely to be enduring the consequences of living with people sick from AIDS in this region alone.

Along with other studies, the data we have collected over the past seven years (for the four-year Orphan Resilience Study and subsequent projects) show that AIDS in a family has major and long-term impacts on children's development and psychological health.

In the Orphan Resilience Study, my team and I asked children whether they were experiencing symptoms such as flashbacks and nightmares. Their responses revealed that children orphaned by AIDS were 117% more likely to be suffering from post-traumatic stress disorder than children whose parents were alive, and also — to our surprise — 67% more likely than children orphaned by other causes, including homicide, suicide and cancer². Researchers in Tanzania, Uganda, Ethiopia, Zimbabwe and China have similarly found heightened psychological distress among children orphaned by AIDS³.

Moreover, our larger, national study in South Africa suggests that 'orphanhood' starts long before a child's parents die. Since 2009, we have ▶

➔ **NATURE.COM**

See *Nature's* vaccine special at: nature.com/vaccines

► interviewed 6,000 children and 2,600 of their caregivers (not necessarily their parents), in six urban and rural sites, as part of the National Young Carers Study (www.youngcarers.org.za). Our preliminary findings indicate that children with caregivers sick from AIDS are just as likely — if not more likely — to suffer from a lasting psychological disorder (depression, anxiety or post-traumatic stress disorder), and just as severely, as children orphaned by the disease (see ‘Cause for distress’). Moreover, 50% more children with caregivers affected by AIDS are afflicted with these disorders than those whose caregivers have other chronic illnesses.

A combination of social factors seems to make having AIDS in a family worse for children than having a death or chronic illness due to other causes. People gossip about the family; the children may be bullied or excluded from the community; and infected caregivers are often severely impoverished and depressed⁴. “They say that my mother is a prostitute and I will die just like her,” one 10-year-old girl told us⁵.

LOSING OUT

Such high levels of psychological distress have severe knock-on effects. One is on education. Several studies, including an analysis of demographic and health-survey data collected by governments in ten sub-Saharan African countries⁶, have shown that orphans do not attend school as much as children with living parents.

Our findings paint a similar picture. In a pilot study, we interviewed 850 children, asking questions such as: “How’s school?” Forty-three per cent of children living with someone ill with AIDS said they couldn’t concentrate due to worry about the sick person at home: one 14-year-old boy said, “I can’t stop thinking about my mother. She looks like she is going to die like my father.” Forty-one per cent missed days at school or had dropped out entirely to care for adults at home.

Children may also pick up infections through toileting the sick person, washing wounds or cleaning soiled bedclothes. As part of the National Young Carers Study, we surveyed more than 5,000 children and caregivers in the KwaZulu-Natal province. Among children living with healthy adults, 4% had symptoms of pulmonary tuberculosis. For those whose caregivers were infected with HIV, this rose to 17% (ref. 7).

Other knock-on effects are higher levels of physical and emotional abuse. Preliminary unpublished findings from the Orphan Resilience Study indicate that 5% of children in healthy families are physically abused

(slapped, punched or hit with a sharp object at least once a week) and 8% are emotionally abused (told at least once a week that they are lazy, stupid, or threatened to be sent out of the house or cursed by an evil spirit). For children living with a caregiver who is sick with AIDS, the numbers rise to 12% and 23%, respectively.

Girls in families affected by AIDS are particularly likely to engage in sex in exchange for money, school fees, transport or shelter. Newly analysed data from the Orphan Resilience Study indicate that girls between 15 and 24 years old in healthy families have a 2.8% chance of being exploited in transactional sex. This climbs to 19% for those with carers sick from AIDS. Among girls with AIDS-sick carers who also experience physical or emotional abuse, 46% say that they have had transactional sex.

The long-term effects of these depredations have yet to be quantified. But various studies worldwide show that school attendance correlates with a child’s chances of getting a job later on. Abuse during childhood is associated with cognitive and social changes that persist long into adulthood — for instance, a smaller hippocampus and reduced capacity for learning and memory⁸. Certainly, the data from our Orphan Resilience Study indicate that psychological disorders worsen as children orphaned by AIDS become young adults, whereas for other orphans and children whose parents are alive, they remain stable.

What can be done to lessen the devastating effects of the AIDS epidemic on upcoming generations?

Happily, the South African government is enthusiastic about implementing research-guided policies for children. The South African government’s departments of social development, health, basic education and agriculture are incorporating the results from our studies and others into national planning policies, such as the 2009–2012 National Action Plan for Orphans and

Other Children made Vulnerable by HIV and AIDS.

The Department of Social Development is also piloting programmes involving life-skills training for affected children and the training of community care workers. For instance, care workers are being taught how to help families with ‘succession planning’ — ensuring that children will be cared for after their parents have died.

INTERVENTION AND SUPPORT

Non-governmental organizations such as the Regional Psychosocial Support Initiative, headquartered in Randburg, South Africa, which serves 13 countries in east and southern Africa, are creating training and support programmes for children living in homes affected by AIDS, based on the results of studies in South Africa, Kenya and Tanzania. These programmes include workshops to teach children caring for sick adults how to lower their risk of infections.

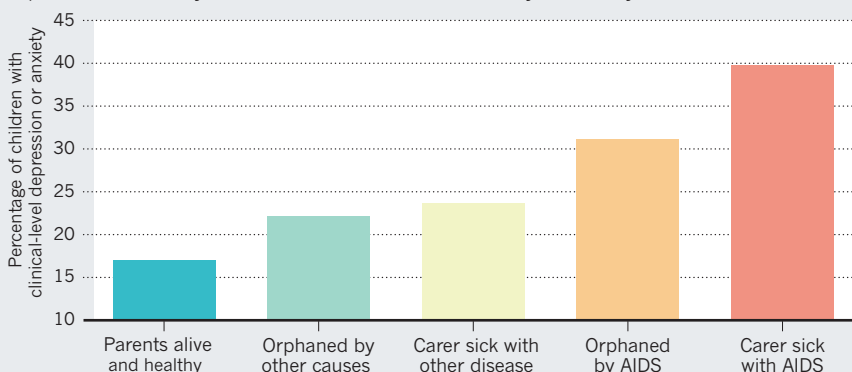
Meanwhile, major development organizations such as the Swedish International Development Cooperation Agency and UNICEF are using and commissioning research to design and improve interventions, including schemes that provide families with money or food.

Yet there is far more to be done. Children’s access to tuberculosis testing must be improved. Also, more parents infected with HIV should be taking antiretroviral medication: a study in Kenya, published by US economists in 2009, showed that children were better fed and had better attendance at school when their parents, unwell with AIDS, started taking antiretrovirals⁹ (see page 29).

Research is needed to understand how to reduce child abuse, and how to make it easier for children to attend and re-enter school — whether by helping them catch up on missed studies, dropping school fees or providing free school meals. Lastly, evidence-based psychosocial interventions, such as cognitive behavioural therapy and support groups,

CAUSE FOR DISTRESS

Children orphaned by AIDS, or whose carers are sick with AIDS, are more likely to have depression and anxiety than children whose lives are similarly affected by other diseases.



SOURCE: ORPHAN RESILIENCE STUDY

are urgently needed for children orphaned by AIDS or living with sick adults.

For such interventions to be effective and accepted, they must be developed through the collaboration of researchers, governments, communities and affected families. So scientists need to focus on the messy, multifaceted social context of the epidemic, not just on the disease itself.

Support programmes must also be properly resourced. Fortunately, organizations such as Save the Children, UNICEF and USAID are already partnering with the governments of South Africa, Swaziland, Malawi and Ethiopia to ensure that essential psychosocial interventions are provided. These include training community care workers in bereavement support.

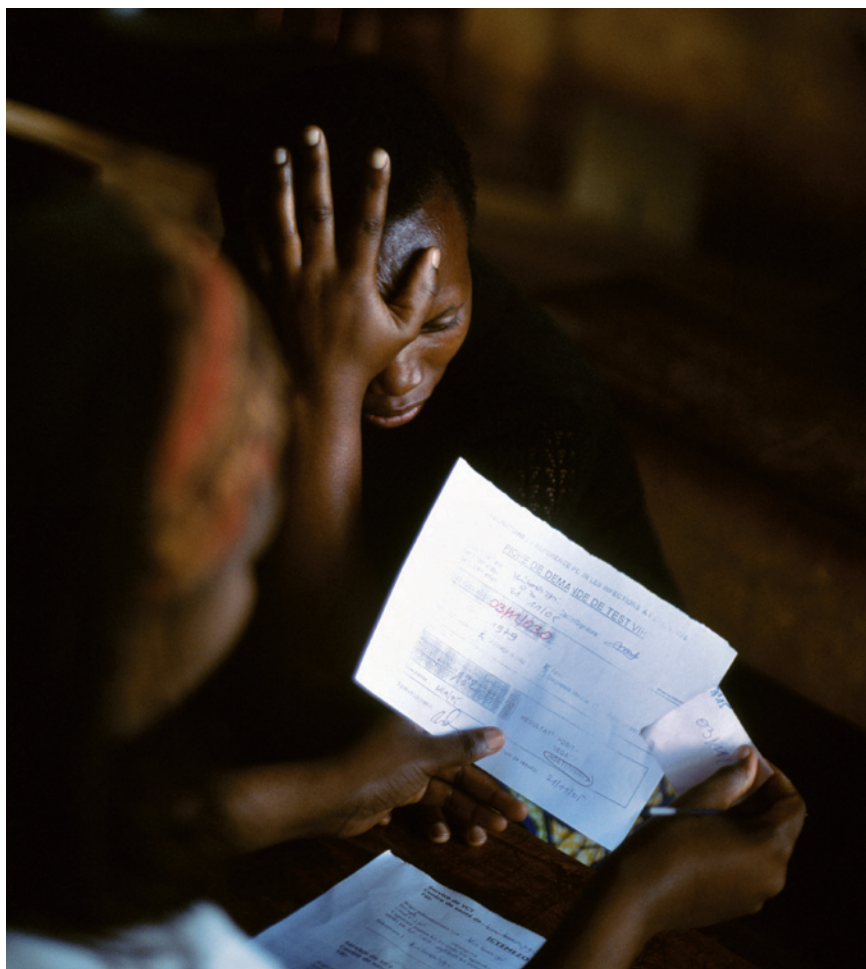
The evidence as to which interventions are effective is still thin, but we are not starting from scratch. Last year, for instance, a collaborative study involving health practitioners from Uganda and Sweden showed that depression, anger and anxiety lessened in children who had lost one or both parents to AIDS if they had access to support groups and intensive medical care¹⁰.

Lindiwe has told only her brother and us about her diagnosis. She asks whether she could call us if she needs to talk to someone. What Lindiwe craves is affection, acceptance. Her boyfriend loves her but in her township, love is Russian roulette: HIV prevalence in her age group is 25%.

A few days after talking to Lindiwe, I spoke in a government meeting in Johannesburg. I put up a graph showing how non-stigmatized children with enough to eat have a 19% risk of developing a clinical-level psychological disorder; children who are stigmatized and hungry have an 83% risk⁴. I looked up at the audience of senior officials. A few of them had tears in their eyes. We must win this battle. ■

Lucie Cluver is at the Department of Social Policy and Intervention, University of Oxford, Oxford OX1 2ER, UK.
e-mail: lucie.cluver@spi.ox.ac.uk

1. Richter, L., Foster, G. & Sherr, L. *Where the Heart Is* (Bernard Van Leer Foundation, 2006).
2. Cluver, L., Gardner, F. & Operario, D. *J. Child Psychol. Psychiatry* **48**, 755–763 (2007).
3. Nyamukapa, C. *et al. Am. J. Public Health* **98**, 133–141 (2008).
4. Cluver, L. & Orkin, M. *Soc. Sci. Med.* **69**, 1186–1193 (2009).
5. Cluver, L. & Gardner, F. *AIDS Care* **19**, 318–325 (2007).
6. Case, A., Paxson, C. & Ableidinger, J. *Demography* **41**, 483–508 (2002).
7. Cluver, L. *et al. Infectious disease and TB co-occurrence amongst children with AIDS-affected caregivers*. Presented at 5th South African AIDS Conference, Durban (2011).
8. Bremner, J. D. *et al. Am. J. Psychiatry* **160**, 924–932 (2003).
9. Graff Zivin, J., Thirumurthy, H. & Goldstein, M. *J. Public Econ.* **93**, 1008–1015 (2009).
10. Kumakech, E., Cantor-Graae, E. & Maling, S. *Soc. Sci. Med.* **68**, 1038–1043 (2009).



A woman receiving her AIDS test result at Kibayi Health Centre in Rwanda.

Stigma impedes AIDS prevention

Medical advances cannot help those who deny they are at risk of HIV and avoid HIV tests. **Salim S. Abdool Karim** describes how such attitudes may be overcome.

Thirty years since the first cases of AIDS were described, there is much to celebrate regarding progress in the treatment and prevention of the disease. Within the past year alone, several studies have revealed that anti-retroviral drugs can prevent the sexual transmission of HIV.

Yet worldwide, many people who are potentially exposed to the virus avoid finding out whether they carry it, or deny that they are at risk of contracting it. Unless people establish whether they are infected, they will not be able to adopt the most

appropriate preventive measures. As scientists and clinicians, our ability to overcome this denialism will determine whether we ultimately succeed in using combinations of all the preventive and therapeutic tools now available to slow, and eventually stop, the HIV/AIDS pandemic.

During its first decade, the disease brought death, pain and suffering, made worse by the stigma that accompanied infection. As a newly qualified doctor in South Africa, it was agonizing for me to watch hundreds of patients die from AIDS — unable to do much more than treat ▶

► their tuberculosis or pneumonia infections and make their last days comfortable. Amid this despair, charlatans and 'snake-oil' cures were plentiful. The real hope among the medical community was that science would produce a vaccine or a cure.

This hope was not entirely misplaced. The most significant advance in the second decade of the epidemic was the development of antiretroviral drugs to treat AIDS patients and to prevent pregnant women from passing on HIV to their children. In the late 1990s, 'triple' antiretroviral therapy transformed AIDS from a uniformly fatal condition to a treatable chronic disease — but only in those countries that could afford the roughly US\$20,000 a year that it cost to treat each patient.

For Africa, the turning point came in 2000, at the 13th International AIDS Conference in Durban, South Africa, when community groups, activists, patients, scientists and health-care workers jointly protested against drug companies and governments, and called for an end to global inequities in access to AIDS treatment. Within a few years, antiretroviral therapy became a reality throughout most of the continent, mainly thanks to funding from the Global Fund to Fight AIDS, Tuberculosis and Malaria, and the US President's Emergency Plan for AIDS Relief (PEPFAR).

DENYING THE RISK

Events took a different course in South Africa, where today 5.4 million people are thought to be infected (see graphic) — more than in any other country. Until 2004, government hospitals were barred from giving antiretrovirals — whether to prevent rape victims from acquiring HIV or to treat patients dying from AIDS. Thabo Mbeki, South Africa's president at the time, denied that AIDS existed. He and his supporters saw the assertion that a new sexually transmitted disease was sweeping through Africa as a racially motivated attack against blacks and a way for pharmaceutical companies to exploit the poor.

Following persistent activism, advocacy, litigation and mass mobilization involving people living with HIV, the Mbeki government eventually initiated a national antiretroviral rollout in 2004. The delay caused an estimated 330,000 unnecessary deaths and some 35,000 avoidable HIV infections in babies¹.

Back in 1999, I was asked to join Mbeki's advisory panel, established to assess whether HIV causes AIDS. Sitting on that panel, engaging in futile debate with AIDS denialists while the epidemic raged on, marked the

lowest point in my life. I never imagined that a decade later, long after Mbeki had stepped down, one of the biggest obstacles to curbing the epidemic, both in South Africa and throughout the world, would be another form of denial — this time at the individual and community level. The two kinds of denialism are very different. But both illustrate that scientific knowledge and innovation alone are not enough to save lives.

In the epidemic's third decade, compelling evidence accumulated from clinical trials that male circumcision reduces the likelihood of men contracting HIV from infected women by more than 50%. Last year, my colleagues and I showed that an antiretroviral drug called tenofovir, formulated as a vaginal gel, reduces the chance of heterosexual women contracting HIV by 39% (ref. 2). Also in 2010, epidemiologists showed that a combination of oral tenofovir and another oral antiretroviral, emtricitabine, reduces the likelihood of men who have sex with men acquiring HIV by 44% (ref. 3). Most promising of all, the results of a multinational clinical trial released last month suggest that antiretroviral treatment can reduce the transmission of HIV from infected men and women to their uninfected partners by 96%.

With these tools — circumcision plus antiretroviral drugs that can treat AIDS patients, prevent mother-to-child transmission of HIV and block the sexual transmission of the virus — stopping the epidemic is within our grasp. But only if more people acknowledge their risk of contracting HIV and find out whether they carry the virus.

Acknowledging infection risk is a first step to getting tested for HIV or taking precautions. Risk awareness must be sustained, as taking preventive antiretroviral pills or applying the microbicidal gel inconsistently can lessen their effectiveness. Also, it is crucial that people who do take preventive measures know whether they already carry HIV, because exposing infected individuals to a single antiretroviral drug designed to prevent uninfected people from contracting the virus could facilitate the emergence of drug resistance.

Moreover, when people know they have HIV, they may be less likely to pass on the virus to uninfected partners. For instance, American men and women who are unaware that they carry HIV are 3.5 times more likely to transmit the virus to others than those who know they are infected⁴.

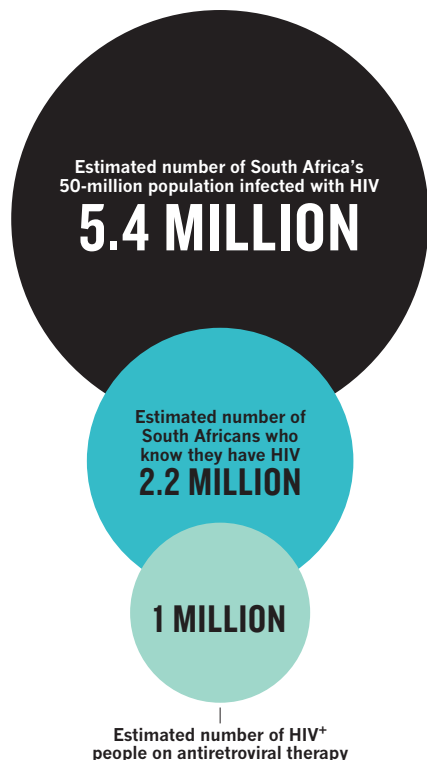
GETTING TESTED

Ensuring that people are sufficiently aware of their level of risk and that they seek regular testing for HIV has proved difficult the world over.

In South Africa, the most recent national HIV survey revealed that 74% of those most at risk of acquiring the virus (including women aged 20–34) were unaware of their HIV status⁵, even though their answers to survey questions indicated that they were well informed about the disease. In 2008, the United States Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, surveyed more than 8,000 men who have sex with men, and found that more than 40% of those infected did not know it⁶. Relatively few of those at risk who are uninfected are taking antiretroviral drugs as a preventive measure⁷, which the CDC recommends.

The numbers volunteering for HIV tests paint a similar picture. Although testing has increased significantly in sub-Saharan Africa in recent years, only about 20% of men and 28% of women in South Africa received an HIV test and result within the past year. Similarly, in Kenya, the 2007 national AIDS survey of 18,000 individuals from nearly 10,000 households found that only 17% of those testing HIV positive reported knowing that they were infected⁸. China fares a little better: about 44% of people living with HIV know that they are infected⁹.

Denial is not the only factor hampering HIV testing and the use of antiretroviral drugs. For instance, in South Africa the public health-care system is already overwhelmed by tens of thousands of AIDS patients. Long queues and overstretched staff discourage many from trying to obtain medical help. But in countries ranging from France to India the practical problems hampering testing and uptake of antiretrovirals are compounded by people seeking care only after they have developed the symptoms of full-blown AIDS.





Women attending an HIV clinic in Rwanda avoid the camera.

L. TOWELL/MAGNUM

We must not allow the fourth decade of the AIDS epidemic to be the decade of missed opportunity. Every prevention programme should involve a combination of interventions tailored to the risks and vulnerabilities of the people receiving them. In sub-Saharan Africa, for instance, teenage girls are up to eight times more likely to contract HIV than boys of the same age. For them, a microbicide gel that they can control is likely to be critical to reducing their risk of infection. Meanwhile, male circumcision is likely to have most impact for men in their early twenties.

The challenge is how to implement these strategies effectively with the limited funds available, so that the successes emerging from clinical trials translate into real-world benefits. The recent call by the CDC and the US National Institutes of Health for proposals for 'implementation science' programmes to address the global epidemic is a step in the right direction. Together, these agencies are committing about \$50 million from PEPFAR to programmes involving interventions known to be effective in clinical trials, and which in combination are likely to improve HIV prevention, treatment and care in poor communities.

Most important, any programme of biomedical interventions should include strategies aimed at encouraging people to acknowledge their risk of contracting HIV.

WORKING TOGETHER

We cannot assume that demand for anti-retroviral drugs, or for any other prevention technology, exists just because the need does. A government attempt to introduce female condoms in South Africa in 1995 illustrated the pitfalls of rolling something out without

concomitantly creating a demand for it. The \$2.6-million effort to distribute 1.3 million condoms to women proved ineffectual because patients did not request them and health-service providers did not adequately promote them.

South Africa's era of denialism at the highest political level illustrated that simply giving people facts and information is not enough. Fear and avoidance must be understood in the context of both the individual and society. Health practitioners and researchers worldwide must engage local communities to find out what factors are preventing people from making the best choices in terms of prevention and treatment. Community members and patients are best placed to advise practitioners how to couch advice and information in ways that are likely to be effective.

This approach has already proved successful in some cases. For instance, my co-workers and I designed a dosing strategy for the tenofovir microbicide gel whereby women apply one dose of gel within 12 hours before sex and a second as soon as possible within 12 hours after sex. We selected this strategy partly on the basis of scientific information about the drug and its effectiveness in animal studies — but also as a result of detailed consultations with rural women in South Africa revealing that having some flexibility over when to apply the gel was essential¹⁰. With this dosing approach, nearly 900 women enrolled in a clinical trial were able to successfully apply the gel about 80% of the time².

An underlying obstacle to finding effective ways to intervene is the separation between biomedical and behavioural research in HIV/AIDS. This emanates

not only from our failure, as researchers, funders and clinicians, to fully appreciate that every biomedical prevention strategy includes a behavioural change, but also from counterproductive hierarchies and territorialism within science. If behavioural and biomedical scientists work together to develop solutions, the coming decade may prove to be the one when the tide was turned against the global AIDS epidemic. ■

Salim S. Abdool Karim is pro vice-chancellor (research) and director at the Centre for the AIDS Programme of Research in South Africa (CAPRISA), University of KwaZulu-Natal, Durban 4001, South Africa, and is in the Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York 10032, USA.
e-mail: karims1@ukzn.ac.za

1. Chigwedere, P., Seage, G. R. III, Gruskin, S., Lee, T. H. & Essex, M. J. *Acquir. Immune Defic. Syndr.* **49**, 410–415 (2008).
2. Abdool Karim, Q. *et al. Science* **329**, 1168–1174 (2010).
3. Grant, R. M. *et al. N. Engl. J. Med.* **363**, 2587–2599 (2010).
4. Marks, G., Crepaz, N. & Janssen, R. S. *AIDS* **20**, 1447–1450 (2006).
5. Shisana, O. *et al. South African National HIV Prevalence, Incidence, Behaviour and Communication Survey 2008: A Turning Tide Among Teenagers?* (HSRC Press, 2009).
6. Centers for Disease Control and Prevention *Morb. Mortal. Wkly Rep.* **59**, 1201–1207 (2010).
7. Bloomberg report; available online at <http://go.nature.com/8wxvi8>
8. *AIDS Epidemic Update 2009* (UNAIDS & WHO, 2009).
9. Zhang, F. *et al. Lancet Infect. Dis.* doi:10.1016/S1473-3099(11)70097-4 (2011).
10. Abdool Karim, S. S. & Abdool Karim, Q. *Nature* **449**, 24 (2007).



The Google company — headed by Sergey Brin (left), Eric Schmidt (centre) and Larry Page — encourages employees to spend time toying with their own ideas.

ECONOMICS

Trial and error

An impressive book argues that we should embrace failure in economic and social progress, finds **Matt Ridley**.

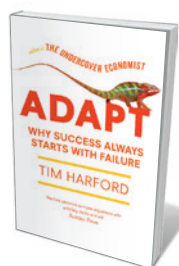
Charles Darwin's big idea — that blind trial and error can progressively build a powerful simulacrum of purposeful design — got pigeonholed under biology. Yet it always had wider implications in economics, technology and culture. Darwin probably drew some elements of his bottom-up thinking from the political philosophers of the Scottish enlightenment, notably Adam Smith and Adam Ferguson. Biology is now returning the favour.

Books such as Steven Johnson's *Where Good Ideas Come From* (Allen Lane, 2010), Kevin Kelly's *What Technology Wants* (Viking, 2010) and Brian Arthur's *The Nature of Technology* (Free Press, 2009) are suffused with concepts from natural selection, as is my own, *The Rational Optimist* (Fourth Estate, 2010). Tim Harford's *Adapt* follows this tradition, focusing on the key role of failure — the 'error' in trial and error — in economic and social progress.

Harford's thesis is that "trial and error

is a tremendously powerful process for solving problems in a complex world, while expert leadership is not". Whether designing computer games, improving foreign aid or discovering how to knock out genes, the heroes and heroines of Harford's book get results not by designing clever solutions and imposing them, but by trying variations and selecting the few that work from among the many that don't. Intelligent design is just as bad at explaining politics and business as it is at explaining evolution.

Harford's case histories are well chosen and artfully told, making the book a delight



Adapt: Why Success Always Starts with Failure
TIM HARFORD
Little, Brown/Farrar, Straus & Giroux: 2011.
320 pp. £20/\$27

to read. But its value is greater than that. Strand by strand, it weaves the stories into a philosophical web that is neat, fascinating and brilliant. Like the best popular science, it advances the subject as well as conveying it, drawing intriguing conclusions about how to run companies, armies and research labs.

The book's message will be music to the ears of many scientists, for Harford exposes the dismal inefficiency of the preconceived, top-down grant-giving that funds much of modern academic research. He celebrates instead the power of prizes and blue-sky funding, and even molecular geneticist Mario Capecchi's documented Nobel-prize-winning decision to use grant money given for one purpose for another. Innovation and discovery come from pluralism and serendipity, not command and control.

Yet this will also be an uncomfortable book for some scientists who read it carefully. For however much they celebrate bottom-up, emergent, evolutionary order in the genome or an ecosystem, most scientists embrace intelligent design as soon as they turn to politics or economics, with government planning playing the part of God. The messy, competitive, pluralistic, unplanned nature of the marketplace is too often anathema to the scientific mind.

A good example is climate policy. Harford shows how exhortations from on high

NATURE.COM
For a review of
Matt Ridley's book:
go.nature.com/8bggnr

D. STRICK/REDUX/EYEVINE

to people to cut their carbon footprints, or winner-picking by governments for advancing certain technologies, is ineffective and counterproductive. Why? Because, he explains (citing chemist Leslie Orgel), “evolution is smarter than we are, and economic evolution tends to outsmart the rules we erect to guide it”. A planning rule that forces British developers to install a minimum amount of on-site energy generation in new office buildings has led to the lunatic spectacle of convoys of diesel-drinking trucks taking carbon-rich wood from forests to biomass boilers in city centres because solar and wind power cannot meet the requirements on such small scales.

Trial and error cannot be used for everything. Nuclear power stations and banks must work without melting down lots of times first. Harford’s analysis of what went wrong at the Piper Alpha oil-rig explosion in the North Sea, the Three Mile Island nuclear meltdown in Pennsylvania and the Lehman Brothers bankruptcy in the United States is illuminating and intelligent. Ineffective safety systems, latent errors and overlooked whistle-blowers are common in all such disasters, but the key ingredient is tight coupling: the systems were designed so that if one part failed, others went with it. To eliminate errors, writes Harford, is “an impossible dream. The alternative is to try to simplify and decouple these high-risk systems as much as is feasible.”

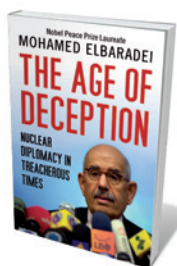
Harford provides some evidence that a new era of evolutionary business is dawning, although the basic idea is as old as the limited-liability company — “a safe space within which to fail”. Companies such as Google have taken the model of the ‘skunk works’ — the name for the trial-and-error division of Lockheed that came up with aircraft such as the U2, the blackbird and the stealth bomber — and rolled it out through the whole firm, by encouraging employees to spend 20% of their time on their own projects. Google’s products, as well as its ideas, are designed so that they evolve by trial and error.

Yet vast swathes of the world are resistant to the implications of this selective approach. Government agencies, academic institutions and financial behemoths are not allowed to fail. “Government regulations,” writes Harford drily, “by their very nature, tend to be somewhat impervious to the possibility of improvement.”

It would be hard to improve Harford’s outstanding book. If pressed, I might say that the focus on variation and selection leaves no room for discussion of the other elements of evolution, especially replication and recombination. Nonetheless, *Adapt* is fine, funny and fluent. ■

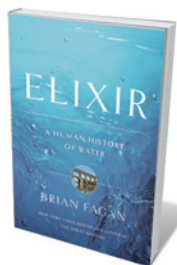
Matt Ridley is author of *The Rational Optimist: How Prosperity Evolves*.
e-mail: mridley@gmail.com

Books in brief



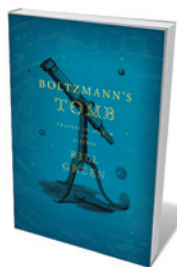
The Age of Deception: Nuclear Diplomacy in Treacherous Times
Mohamed ElBaradei BLOOMSBURY 352 pp. £20 (2011)

Mohamed ElBaradei served as director-general of the International Atomic Energy Agency from 1997 to 2009, and shared the Nobel Peace Prize with the agency in 2005. In his memoir, he reflects on the use of diplomacy to limit nuclear proliferation across the globe, giving an insight into foreign policy approaches. He charts his experience of negotiations during conflicts with North Korea and Iran, and in the run-up to the Iraq war, highlighting the difficulty of maintaining objectivity when under pressure from presidents, politicians and the press.



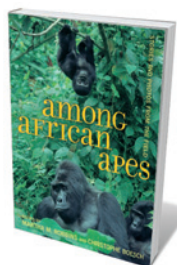
Elixir: A Human History of Water
Brian Fagan BLOOMSBURY 416 pp. £20 (2011)

The availability of water is central to human survival and the growth of civilization. Anthropologist Brian Fagan charts three ages of water in his book. In the first, water was scarce and it was worshipped by early humans. In the second, we learned to manage water through engineering — notably the innovations of the Greeks, Romans and Victorians. We are now entering the third age, he contends, when we will again have to learn to revere this essential liquid, albeit with the benefit of improved technology. Lessons from the past could help us adapt to a drier future, he suggests.



Boltzmann's Tomb: Travels in Search of Science
Bill Green BELLEVUE LITERARY PRESS 288 pp. \$25 (2011)

In a history of science told as a travelogue, geochemist Bill Green describes his pilgrimage to the places in which his scientific heroes worked and made groundbreaking discoveries. He hunts down the remnants of Antoine Lavoisier’s chemistry laboratory in Paris, finds signs of astronomer Johannes Kepler in Prague, visits Albert Einstein’s apartment in Bern and locates Boltzmann’s tomb in Vienna, relating tales of many others along the way. The result is a very personal take on science’s twists and turns throughout the centuries.



Among African Apes: Stories and Photos from the Field
Martha M. Robbins and Christophe Boesch UNIVERSITY OF CALIFORNIA PRESS 200 pp. \$29.95 (2011)

By relating first-hand stories of their field work with wild African great apes, primatologists Martha Robbins and Christophe Boesch give a fresh insight into research and conservation efforts. Their vivid descriptions reveal how chimpanzees, bonobos and gorillas hunt, socialize and play in their natural habitat, as well as the threats they face from poaching, disease and deforestation. The authors also describe their experiences of working with the animals in remote places, and highlight the effectiveness of conservation efforts.



Between Raphael and Galileo: Mutio Oddi and the Mathematical Culture of Late Renaissance Italy

Alexander Marr UNIVERSITY OF CHICAGO PRESS 384 pp. \$45 (2011)
Mutio Oddi of Urbino is less famous than his contemporary Galileo Galilei, but made many contributions to mathematics, instrument-making and architecture in the seventeenth century. Art historian Alexander Marr delves into Oddi’s archives to piece together the typical life of an artisan-scholar in late Renaissance Italy. He shows how scientific advances then depended more on who you knew and where you lived than on the breakthrough you achieved.

NEUROSCIENCE

Recalling the future

Our memories may be tools for working out what's to come and what to do about it, finds **Linnaea Ostroff**.

Memory is the most intimate of abstractions. The matter of how memories are made and stored is a research goal that needs no selling, perhaps driven by a fear of memory loss and our dread of personal obliteration. Yet one question has been neglected: why does memory exist?

Possible answers are explored in *Predictions in the Brain*, a collection of 25 rigorous, data-laden cognitive-science reviews edited by neuroscientist Moshe Bar. He and his co-authors propose that prediction is a unifying principle of brain function, and that predictions are created from memories. As contributor Yadin Dudai writes: "Memories are made mostly for the sake of present and future." Memory systems do not store past experiences, but recycle their components into the imagined future.

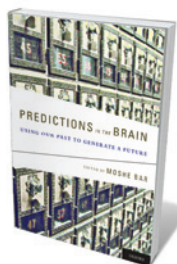
In throwing evolutionary light on a fundamental process, this idea has legs. As survival advantages go, our ability to envision and plan a nuanced future is a masterpiece — arguably, the root of our success as a species. If memories are used to generate predictions that drive our actions, then the mechanisms of prediction are as worthy of study as those of memory. The growing body

NATURE.COM
How to win memory competitions:
go.nature.com/chab64

of experimental data reviewed in this volume supports a shift in memory research from storage to prediction.

From anticipating fine motor movements to imagining our old age, what we think will happen in the future depends on what we know has been. We hit a baseball on the basis of how we have done it before; our experience of traffic lights allows us to anticipate and halt in front of the red one. With only past events to rely on, however, accurate prediction is difficult. A memory system must store information reliably, yet be flexible enough to keep knowledge up to date and to know which items are relevant to a particular purpose. The brain manages these challenges well. It is easy to think of memories as faithful, discrete records of our past, pressed and dried like items in a scrapbook. They are not — and for good reason.

Dudai explores how recalling memories



Predictions in the Brain: Using Our Past to Generate a Future
EDITED BY MOSHE BAR
Oxford University Press: 2011. 400 pp.
\$99.95

makes them plastic, susceptible to updating or the addition of new associations. This mechanism serves prediction by providing us with the most up-to-date information. For example, if the train to work has been on time 4 out of 5 days, taking it again would seem to be a good decision. But if by next month this has become 6 out of 20 days, choosing the train looks unwise. Each memory of the train must be associated with the previous and subsequent ones as they occur, to estimate the chance of it being late tomorrow.

Beyond the constant generation of everyday predictions (I must push the door to open it; the milk will run out soon), our brains also create detailed pictures of future events. This capacity is thought to be uniquely human, and occupies a significant portion of our mental activity. It is entwined with episodic memory — our personal collection of autobiographical clips. *Predictions in the Brain* reviews experimental evidence for substantial overlap between memory recall and future simulation. Generally, patients with deficits in episodic memory are unable to envision a complex future; and healthy subjects show similar brain activation when recalling the past and imagining the future.

Several reviews describe the construction system of the brain, which assembles imagined events from our past knowledge. We create altered versions of remembered situations by recombining primitive images with digested information from memory. Our experience of time may be a consequence of a single, versatile storytelling mechanism in the brain that draws its cast, plot and scenery from the memory banks. As authors Demis Hassabis and Eleanor Maguire point out, time is simply one more component of this drama.

As for the neural mechanisms responsible for imagining the future, several brain systems are thought to cooperate in generating mental constructs. Various cortical areas are implicated, and one suggestion is that the hippocampus combines and integrates information stored in other systems. Through this lens, dreams could be seen as shuffles through the mental primitives.

We have always been uncomfortable with explicit prediction: the mythical Cassandra was hated for her supernatural foresight, the major religions forbid fortune-telling, and we are anxious about testing ourselves for genetic diseases that have a delayed onset, such as Huntington's. But *Predictions in the Brain* shows that this rejection of a certain future is legitimate if much of our mental life is devoted to looking ahead. Memories are not the flotsam of the past: they are the raw materials of cognition. ■

Linnaea Ostroff is a researcher at the Center for Neural Science, New York University, New York 10003, USA.
e-mail: lostroff@nyu.edu



Hitting a baseball requires the brain to predict what might happen next, based on past experience.



Composer Eduardo Reck Miranda (back left, at computer) synthesizes music from human voice sounds.

Q&A Eduardo Reck Miranda

Musical intelligence

Eduardo Reck Miranda is a composer and leading researcher in artificial intelligence in music, based at the University of Plymouth, UK. A revised version of his *Sacra Conversazione* — five movements for string orchestra, percussion and electronics — will be performed on 9 June in London. He explains what music can tell us about speech, physiology and cognition.

In *Sacra Conversazione* you synthesize 'artificial words' by splicing sounds from different languages. What inspired this?

The composition focuses on the non-semantic communicative power of the human voice, which is conveyed mostly by the melodic contour, rhythm, speed and loudness of vocal sounds. There is evidence that the non-semantic content of speech, such as emotional intent, is processed by the brain faster than semantic content: humans seem to have evolved a 'fast lane' for this non-semantic content. I believe that this aspect of our mind is central to our capacity for making and appreciating music.

How did you create the artificial words?

I started by combining single utterances from several languages. I used more than a dozen — as diverse as Japanese, English,



Spanish, Farsi, Thai and Croatian — to form hundreds of composite 'words', as if I were creating the lexicon for a new artificial language. It was a painstaking job, using sophisticated speech-synthesis methods. Yet I was surprised that only about one in five of these new 'words' sounded natural.

Why didn't the assembled words sound realistic?

The problem was in the transitions between the original segments. For example, the transition from, say, Thai utterance A to Japanese utterance B did not sound right. But the transition of the former to Japanese utterance C was acceptable. I have come to believe that the main reason is physiological. When we speak, our vocal mechanism articulates

Sacra Conversazione
Queen Elizabeth Hall,
London. 9 June 2011
at 7.30 p.m.
On air and online on
BBC Radio 3 on 3 July.

a number of muscles simultaneously. So if we synthesize artificial utterances that are physiologically implausible, the brain is reluctant to accept

them. Human-voice perception — and, I suspect, auditory perception in general — is very much influenced by the physiology of vocal production.

What other approaches did you try?

I tried to synthesize voices using a physical model of the vocal tract. The model has more than 20 variables, each of which roughly represents a particular muscle. But I found it extremely difficult to produce decent utterances with this model. All the same, I used some of these sounds in the composition: they sounded voice-like but not word-like. This explains why artificial speech technology is still so reliant on splicing and smoothing methods.

How did you then turn these sounds into music?

Synthesis and manipulation of voice are only the cogs, nuts and bolts. Music happens when one starts to assemble the machine. It is hard to describe how I composed *Sacra Conversazione*, but inspiration played a big part. Creative inspiration is beyond the ability of computers, yet finding its origin is the holy grail of the neurosciences. How can the brain draw up and execute plans on our behalf implicitly, without telling us?

Is the evolving understanding of music cognition opening up possibilities in music composition?

Yes, to a limited extent. But progress will probably emerge from the reverse: new possibilities in musical composition will contribute to the development of such understanding. Cognitive neuroscience methods force scientists to narrow the concept of music, whereas I want to broaden it. But the approaches are not incompatible: each can inform and complement the other.

What are you working on now?

I am working on a human-computer interface in which the user can control musical parameters solely with the brain (see go.nature.com/ieds9g). And I am orchestrating plots of spiking neurons and the behaviour of artificial-life models for *Sound to Sea*, a large-scale symphonic piece for orchestra, church organ, percussion, choir and mezzo-soprano soloist. The piece will be premiered in 2012.

What do you hope audiences will feel when listening to your work?

My main aim is to compose music that is appreciated as a piece of art rather than as a challenging auditory experiment. If the music makes people think about the relationship between sound and language, I will be even happier. Music is not merely entertainment. ■

INTERVIEW BY PHILIP BALL

CORRESPONDENCE

Respect limits of embryo patents

In what could be a milestone case, the Court of Justice of the European Communities is expected to clarify and define the term human embryo as used in European patent law within the next few months. But patent law should not be driven solely by the interests of patent holders (*Nature* 472, 418; 2011).

The environmental group Greenpeace started this court case to obtain clarification of European patent directive 98/44/EC, which prohibits the granting of patents on “uses of human embryos for industrial or commercial purposes” and on “the human body at the various stages of its formation and development”.

The European patent granted in 1999 to Austin Smith, Peter Mountford and the University of Edinburgh, UK (EP 0695351; 1999), covered human embryos, embryonic stem cells and their production from human embryos. Opposition from Greenpeace and others, including the German and Dutch governments, caused the European Patent Office (EPO) to revoke large parts of the patent.

In 2008, the EPO decided in another precedent case (G2/06) not to grant patents on human embryonic stem cells produced by the destruction of human embryos. The number of European patent applications in stem-cell research has since increased. We expect the European Court of Justice to make a similar decision to the EPO and think that, in general, stem-cell research will not be affected.

We believe that patents should not encourage the production or destruction of human embryos for commercial use. Before the European patent directive came into force in 1998, the scientific community and the biotechnology industry had

lobbied for its approval. Clear ethical limits were a precondition for its adoption by the European Parliament. Science and industry should now respect these limits.

Roland Hipp* *Greenpeace, Germany.*

Peter Liese* *Member of European Parliament, Chairman of the European People's Party Working Group on Bioethics.*

**On behalf of 4 co-signatories. See <http://dx.doi.org/10.1038.474036a> for a full list of signatories and for a declaration of competing interests.*

Cameroon: not ready for REDD+

Contrary to your implication (*Nature* 472, 390; 2011), the World Bank's policy on indigenous peoples depends on more than just their consent after consultation.

The charter of our Forest Carbon Partnership Facility (FCPF) makes it clear that World Bank safeguards, which ensure development finance is environmentally and socially sustainable, apply to FCPF-financed activities. This includes the bank's policy on indigenous peoples: the bank will proceed only with projects supported by affected communities after free, prior and informed consultation.

Regarding Cameroon's REDD+ plan (which goes beyond ‘reducing emissions from deforestation and forest degradation’ to include conservation and sustainability), the FCPF has not yet disbursed any funds to the country as it still has to draft its national REDD+ plan. It is therefore too early to judge the plan's content or the quality of consultations.

So far, Cameroon has submitted to the FCPF only its initial idea note for the Readiness Preparation Proposal, prepared with the environment ministry (not with the conservation group

WWF, as suggested in your Editorial) in 2008. The FCPF reviewed this document and has advised Cameroon how to achieve adequate consultation during the drafting of the next stage of its national proposal.

Cameroon will therefore not be in a position to present its proposal at this month's meeting of the FCPF participants committee.

Benoît Bosquet *FCPF, World Bank, Washington DC, USA.*
bbosquet@worldbank.org

Cameroon: listening to indigenous peoples

A report by the Forest Peoples' Programme (FPP) criticizes the REDD forest-emissions reduction process in Cameroon for its perceived lack of measures to protect and benefit forest communities (*Nature* 472, 390; 2011). It is premature to cry foul at this point as the process has yet to start.

The conservation group WWF and its partners recognize that indigenous and local communities' rights will be strengthened by including them in national REDD programmes, ultimately putting management of their lands back in their hands. Hence a pre-meeting has already been held for indigenous peoples. With its knowledge of social safeguards, the FPP's contribution will be valuable to this process.

The WWF is set to undertake a REDD pilot project in the Ngoyla–Mintom region, where the indigenous Baka people are crucial rights-holders and full participants in determining the future of this 1-million-hectare forest. The WWF believes that such efforts will help to establish successful national REDD programmes and to ensure that the forests on which these communities depend are maintained.

Much work lies ahead to deliver REDD's promises on climate, biodiversity and people's well-being. But it should not be declared a failure before it starts.

David Hoyle *WWF Cameroon.*
djhoyle@wwf.panda.org

Population mapping of poor countries

Global population maps can be valuable for quantifying populations at risk, such as those near nuclear power plants (*Nature* 472, 400–401; 2011). But the uncertainties inherent in such data sets must be acknowledged. The census data used in map construction for rich countries are recent and detailed. The same is often not true for poorer countries.

For example, Angola's last census was in 1970, broken down into just 18 districts. Estimates of its current total resident population vary from 13.3 million to 19 million, according to the US Census Bureau and the United Nations, respectively. When such outdated and coarse-resolution data are subject to different modelling assumptions by different groups, it can lead to substantially divergent estimates of population distributions and, consequently, populations at risk.

Uncertainties in and between global population maps should be more widely discussed, and a greater effort made to quantify them. Furthermore, spatially referenced demographic data used in map construction are often scattered across national statistical offices and websites. A centralized, open-access, up-to-date database would benefit many fields that rely on population maps, and would require minimal investment.

Andrew Tatem *University of Florida, USA.*

andy.tatem@gmail.com

Catherine Linard *University of Oxford, UK.*

Willard Boyle

(1924–2011)

Physicist who helped invent the ‘eye of the digital camera’.

One afternoon in 1969, Willard Boyle and I had a brainstorming session in his office at AT&T Bell Laboratories in Murray Hill, New Jersey. Our aim was to work out how to move electrical charge through a semiconductor to make a memory device for computer applications.

At the time, researchers were pursuing a type of computer memory called ‘bubble memory’. This involved using a magnetic field to create small magnetized bubbles, each of which stored one bit of data, within a thin film of magnetic material, and then making these shift from one site within the film to the next. With an input device to inject the bubbles and a detection device at the other end, the system offered a way to preserve and retrieve information.

Boyle and I knew that we couldn’t create magnetic bubbles in a semiconductor. But we wondered about storing the charge that exists in a semiconductor in a device called a metal-oxide-semiconductor capacitor. Eventually we worked out that by placing two such capacitors very close together, we could make charge pass from one capacitor to the next.

Our theory for how to produce a charge-coupled device (CCD) — versions of which are now found in technologies from digital cameras and bar-code scanners to the Hubble Space Telescope — was completed in about an hour. After we published papers demonstrating proof of principle, more than one person commented, “I could have invented that if I’d thought of it.”

Boyle, who in 1969 was in charge of research and development for semiconductor devices (and my boss), had arrived at Bell Labs via an unusual route. He was raised in a remote logging community in Sanmaur, Quebec, and schooled by his mother. He went to Lower Canada College at the age of 14 and then to McGill University, both in Montreal, for 7 years of formal education.

For a few years, the Second World War pulled him away from his studies. Soon after joining the Royal Canadian Navy in 1943, he served as a Spitfire pilot. Indeed, he later took pride in wearing his ‘Fleet Air Arm’ wings on his lapel. Once the war was over,

he returned to McGill, eventually earning a doctorate in physics in 1950. Three years later he joined Bell Laboratories, initially as a member of the company’s technical staff.

Soon after our brainstorming session in his office, a rapidly assembled team of researchers produced the first crude CCD. We then demonstrated that the CCD could in principle be used as a type of digital circuit — a shift register — and as an imaging device. For the latter, charge could be injected into the system by shining light on the array of capacitors.

Fortuitously, my ‘Device Concepts’ department was developing a silicon diode-array camera tube — a device used to convert

to observe the Universe by several orders of magnitude. Boyle and I shared half of the 2009 Nobel Prize in Physics for our invention. (Charles Kao received the other half for his work on optical fibres.)

Boyle’s other major contributions include the first continuously operating ruby laser, which he invented with Don Nelson in 1962. He was also awarded the first patent (with David Thomas) proposing a semiconductor laser. Although the technology needed to produce the laser wasn’t available when Boyle proposed it, semiconductor lasers are now used in a vast range of applications, including in compact-disc players.

In 1962, Boyle became director of space science and exploratory studies at Bellcomm, an AT&T subsidiary that provided technological support for NASA’s Apollo space programme. While at Bellcomm, he helped NASA researchers select the best spot for man’s first landing on the Moon. In 1964 he returned to Bell Labs and switched from pure research in solid-state physics to developing electronic devices, particularly silicon integrated circuits — the now essential building blocks of telecommunications, computers and electronics in general.

Both as a research scientist and as an administrator, Bill was continually looking for new ideas and new ways of doing things. He

always gave the impression that he was having fun and not just doing a job. Once he demonstrated the feasibility of his concept for a new type of snow-making machine in his garage.

He had an insatiable curiosity and love of life. He loved experimenting with his digital camera, showing his work in local galleries — and with four children (one deceased), eight grandchildren and one great grandchild, his family life was a full one. An intrepid traveller, he and his wife Betty were driving around the Canadian Maritimes in their Mini Cooper — with its ‘CCD’ licence plate — even in the last week of his life. ■

George Smith is based in Barnegat, New Jersey. He worked with Willard Boyle for about ten years at Bell Labs.
e-mail: apogee2@comcast.net



Willard Boyle (left) and George Smith in 1974 with their charge-coupled device.

an optical image into an electrical signal — for use in a ‘Picturephone’. (The Picturephone never took off but the idea was to allow people to see as well as hear each other over the telephone.) The possibility of an imaging device that didn’t rely on bulky, inefficient tube-based cameras prompted a research programme at Bell Labs to develop a CCD video device. The first CCD imaging device was produced in 1970. Other companies, including the Radio Corporation of America, Texas Instruments, Fairchild Camera and Sony soon developed the CCD concept for an array of applications.

Today, the CCD has replaced electron-beam TV cameras and photographic film, and is used in scanning devices, medical imaging and in space applications including in military surveillance satellites. Indeed, the CCD has improved astronomers’ ability

MALARIA

Mosquitoes bamboozled

Mosquitoes harbouring the malaria parasite follow pulses of carbon dioxide exhaled by their human prey to track them down. A novel set of chemicals can distract these insects by disrupting their sense of smell. [SEE LETTER P.87](#)

MARK STOPPER

The mosquito could be the world's most dangerous creature. It is a vampire vector for the malaria parasite and other pathogens that kill about a million humans, mostly children, each year. People protect themselves against mosquitoes by sleeping under fine netting and by slathering themselves with high concentrations of DEET — an effective, if expensive, repellent. But on page 87 of this issue, Turner *et al.*¹ report a promising new line of defence.

Mosquitoes follow their 'noses' to reach their victims². When humans exhale, they release clouds of carbon dioxide. Gusts of wind turn these clouds into plumes consisting of discontinuous dashes and dots of vapour, forming a signal to mosquitoes that spells dinner. Even far from the source, mosquitoes are particularly sensitive to the plume's structure: the on-and-off pulses of CO₂ seem necessary to motivate and direct their upwind flight³ (Fig. 1a).

The nose with which mosquitoes detect CO₂ is a structure called the maxillary palp. There, an olfactory sensory neuron known as cpA detects a short pulse of the gas and responds by generating a characteristic signalling pattern — a quick burst of action potentials followed by a pause⁴. A dedicated pathway then transmits these signals through the mosquito's brain. The proteins and structures involved in this transduction process are highly conserved across a range of species, including several types of mosquitoes and fruitflies. The cpA neuron thus stands as a potential fulcrum for powerful leverage: by interfering with this neuron's ability to respond to the on-and-off CO₂ pulses, it might be possible to disrupt a mosquito's bloodthirsty manhunt.

Turner *et al.*¹ report new ways to do this. Their work began a few years ago with a curious observation in fruitflies: although, ordinarily, CO₂ repels these insects (CO₂ is the main ingredient of a compound released by flies to alert their fellow flies to danger⁵), they are strongly attracted to ripe fruits, which emit large amounts of this gas. Previously, Turner and Ray⁶ found that flies can overcome this repulsion because ripe fruits also release 'inhibitory' odorants that effectively block CO₂ detection. The inhibitory odorants prevented

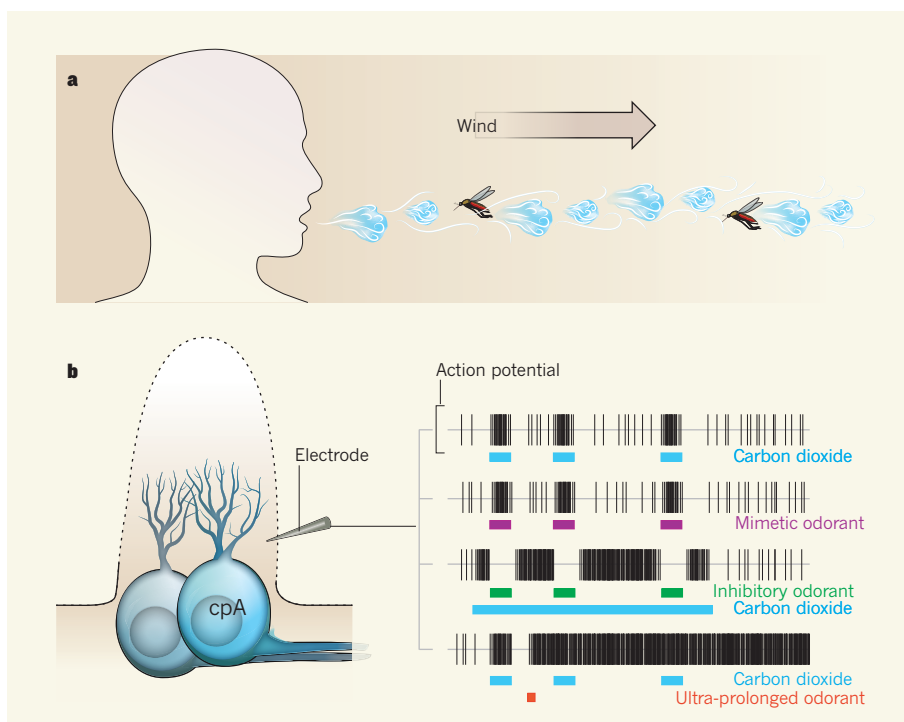


Figure 1 | Newly identified odorants can fool mosquitoes. **a**, Clouds of CO₂ exhaled by humans are segmented by wind into a series of pulses of vapour separated by clean air. Bloodthirsty mosquitoes track humans by following these pulses upwind, a process mediated by a specific olfactory sensory neuron called cpA (**b**). The cpA neuron responds to CO₂ pulses with a brief burst of action potentials. Mimetic odorants such as 2-butanone evoke responses identical to those of CO₂. Inhibitory odorants can silence the cpA neuron's response to CO₂. And ultra-prolonged odorants evoke continuous cpA-neuron firing, rendering it unresponsive to the separate CO₂ pulses that signal human presence.

the cpA neurons in mosquitoes — and their analogues in fruitflies — from firing in the presence of CO₂.

Using the chemical structures of these inhibitory odorants as templates, Turner *et al.*¹ now identify several compounds that also affect the mosquito's cpA neuron (Fig. 1b). Some of these compounds join the list of inhibitory odorants. One compound, 2-butanone, precisely mimics the effect of CO₂, eliciting a brief burst of action potentials followed by a pause. But another, 2,3-butanedione, causes the cpA neuron to do something quite extraordinary: a 1-second pulse of the odorant elicits an 'ultra-prolonged' period of intense firing lasting at least 5 minutes — the full length of the recording session (Fig. 1b).

To elicit ultra-prolonged firing, the authors

formulated a powerful blend of solvent plus odorants, including 2,3-butanedione, that worked together at low doses. The ultra-prolonged activation effect of the blend on the cpA neuron was highly specific, triggering no response from a neighbouring neuron. Moreover, once the blend had activated the cpA neuron, subsequent pulses of CO₂ or other odorants had no effect on the neuron's runaway train of spikes: while firing wildly, the cpA neuron had become insensitive to the on-and-off plumes of CO₂ that beckon mosquitoes to people.

But could this blend actually disorient hungry mosquitoes? To test this, Turner and colleagues released individual mosquitoes into one end of a wind tunnel while blowing air containing a plume of CO₂ from the other end.

Under these controlled conditions, the authors tracked individual mosquitoes in flight.

Insects that had been pre-exposed to solvent alone had no trouble flying upwind along the CO₂ plume to its source. However, mosquitoes pre-exposed to the ultra-prolonged blend seemed to get lost along the way — they were slow to get started, meandered, and took more time to reach the source, if they reached it at all. Because the insects had been pre-exposed to the blend before entering the wind tunnel and did not encounter it during the test, the authors could rule out the possibility that the misdirected mosquitoes had been distracted in flight by detecting blend components through other types of receptor neurons. This result directly links the insensitivity of the cpA neuron to the inability of the mosquito to track CO₂.

For a more realistic test, Turner and colleagues conducted a study in Kenya. There, within a large enclosure, they assembled two

draughty huts typical of the region. Both huts contained mosquito traps releasing alluring plumes of CO₂, but one also contained a vaporizer dispensing the authors' ultra-prolonged blend. The mosquitoes were released into the enclosure overnight. By morning, compared with the relatively unprotected hut, only half as many mosquitoes had been trapped in the hut containing the ultra-prolonged blend.

These results¹ bode well for the hunt to find a means of avoiding mosquitoes. Inhibitory odorants might help to mask the presence of humans; CO₂ mimetics may provide advantages for luring mosquitoes into traps; and ultra-prolonged activation agents dispersed through the air might shield whole groups of people. But because mosquitoes are also attracted to other human body odours in sweat, breath and skin⁷, it remains to be seen how effective these compounds will be for protecting people.

Moreover, as the authors note, the chemicals they have tested so far, including 2,3-butanedione, have not been shown to be safe for humans. But the principles that these compounds reveal are definitely not safe for mosquitoes. ■

Mark Stopfer is at the National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA.
e-mail: stopferm@mail.nih.gov

1. Turner, S. L. *et al.* *Nature* **474**, 87–91 (2011).
2. Gillies, M. T. *Bull. Entomol. Res.* **70**, 525–532 (1980).
3. Geier, M., Bosch, O. J. & Boeckh, J. J. *Exp. Biol.* **202**, 1639–1648 (1999).
4. Lu, T. *et al.* *Curr. Biol.* **17**, 1533–1544 (2007).
5. Suh, G. S. *et al.* *Nature* **431**, 854–859 (2004).
6. Turner, S. L. & Ray, A. *Nature* **461**, 277–281 (2009).
7. Zwiebel, L. J. & Takken, W. *Insect Biochem. Mol. Biol.* **34**, 645–652 (2004).

QUANTUM INFORMATION

Entanglement as elbow grease

Quantum correlations have long been recognized as an informational resource for quantum communication and computation. It now seems that they can also do physical work. SEE LETTER P.61

PATRICK HAYDEN

It is a fact, confirmed through mundane and repetitive experience, that it's much easier to make a mess than to clean one up. Our personal battles to keep our desks tidy and our houses clean remind us every day that nature is inclined to disorder. That preference is so pronounced that it has a name: the second law of thermodynamics. Entropy measures disorder, and the second law states that entropy cannot decrease. On page 61 of this issue, del Rio *et al.*¹ show that, even though the second law itself is inviolable, clever quantum engineering can sometimes make it possible to sneak around the law's well-known consequence that erasing information requires an investment of physical work.

The importance of erasure becomes clear from asking a simple question. Is computation, with its constant and complicated reordering of the states of the computer, more akin to cleaning a house or messing it up? Back in 1961, Rolf Landauer made a profound, if tautological, observation²: the only necessarily irreversible steps in a computation are those that erase information. Erasing a bit of information (which can have a value of either '1' or '0') by setting it to '0' is the logical equivalent

of cleaning a messy house, which means that it should take some effort. Indeed, resetting a bit from an unknown state to '0' reduces the computer's entropy by exactly one bit, so the second law of thermodynamics requires that bit of entropy to get shuttled somewhere else.

Ultimately, the entropy ends up in the environment, in the form of an amount $k_B T \ln 2$ of heat, where k_B is Boltzmann's constant and T is the temperature. Anyone who has noticed their computer's temperature rise precipitously as it works on a demanding problem, such as decoding a streaming video, has witnessed at first hand the dissipation of heat associated with irreversible computation. (Of course, today's technology is thermodynamically inefficient, dissipating orders of magnitude more heat than required by Landauer's principle.)

In fact, there is no fundamental need for computation to erase bits at all. A computer could just use fresh bits at every stage of its computation, never erasing anything. But in reality, such a computer would rapidly run out of memory. In the 1980s, however, Charles Bennett found³ a cleverer approach based not on storing used bits but on reversibly uncomputing them. Unlike erasing bits, this eliminates the need to dissipate heat. Bennett's reversible computation has a central role in quantum computation

because fundamental logic gates in quantum computation are naturally reversible.

So, if inventive physicists and engineers manage to build a quantum computer someday, it will be based on reversible logic, never erasing any of its quantum bits (qubits). Or will it? In their study, del Rio *et al.*¹ present a counter-intuitive revision of Landauer's principle that is applicable to quantum states. In some circumstances, erasing a qubit need not heat up the environment at all, but instead can cool it down. Equivalently, instead of having to invest work to erase the qubit, the process of erasing the qubit can actually generate work, like a tiny quantum-logical wind turbine.

As is often the case with surprising features of quantum information, entanglement — a kind of superstrong quantum-mechanical correlation — is the resource making the effect possible. One of the fundamental lessons of quantum information theory is that entanglement is more than just a conceptual puzzle; it is a resource that can be exploited, for example to send qubits over long distances via teleportation⁴ or to establish secrets for use in cryptography⁵. The study by del Rio *et al.* demonstrates convincingly that entanglement is not just an informational resource, but a physical one. Entanglement can literally do work.

To see how this comes about, consider trying to design a machine that will erase a bit, resetting it to '0'. If the bit is already '0', the machine doesn't need to do anything. If the bit is '1', the machine needs to flip it. Neither of these actions is irreversible, so it would seem possible to erase the bit without leaking any heat to the environment, violating Landauer's principle. The trick is that, to decide whether to leave the bit alone or to flip it, the machine has first to read the bit, inadvertently creating a copy of it somewhere else. Such a machine isn't erasing the bit so much as just moving it around in memory. The way to avoid such sneaky sleights

of hand is to keep track of not just the bit being erased but the whole computer memory, requiring that the memory be left just as it was found once the erasure process has finished.

Using the symbol S for the system being erased and M for the memory, the amount of work required to erase the system is $H(S|M)k_B T \ln 2$. The function $H(S|M)$, known as the conditional entropy of S given M , measures how much residual uncertainty about S there is, given the contents of M . If S is a pair of random bits, for example, M might store whether those two bits are the same or different. In that case, although S itself contains two bits of entropy, $H(S|M) = 1$ because, given M , knowledge of one bit is enough to figure out the other, leaving only one bit of residual uncertainty about S .

As usual, though, quantum mechanics makes things a bit stranger. If the system and memory are entangled with each other, then $H(S|M)$ can be negative: a machine with access to the memory M can be “more than certain” about the system S (ref. 6; for a non-technical discussion, see ref. 7). Mathematically, the reason this happens is that the conditional entropy can be expressed as $H(S|M) = H(S, M) - H(M)$, the amount of entropy in the pair (S, M) that cannot be attributed to M . Highly entangled states can have high-entropy parts and a low-entropy whole — that is, large $H(M)$ but a small $H(S, M)$, leading to a negative $H(S|M)$. Intuitively, this is a reflection of the fact that there is more to an entangled quantum state than just the assemblage of its disordered parts. In their study, del Rio and colleagues prove that the work cost of erasure in quantum mechanics remains $H(S|M)k_B T \ln 2$, even when $H(S|M)$ is negative, in which case erasing doesn't cost work, it generates it.

The authors¹ explain how to build a machine that would extract the work, but other than in the simplest cases, the machine would have to execute some sophisticated quantum computations. No ordinary heat engine, its building blocks would be highly efficient quantum error-correcting codes first invented to solve communications problems. Del Rio *et al.* creatively adapt those codes to establish a new and fundamental link between the information content of quantum correlations and more familiar physical concepts such as heat and work. ■

Patrick Hayden is in the School of Computer Science, McGill University, Montreal, Quebec H3A 2A7, Canada.
e-mail: patrick@cs.mcgill.ca

- del Rio, L., Åberg, J., Renner, R., Dahlsten, O. & Vedral, V. *Nature* **474**, 61–63 (2011).
- Landauer, R. *IBM J. Res. Dev.* **5**, 183–191 (1961).
- Bennett, C. H. *SIAM J. Comput.* **18**, 766–776 (1989).
- Bennett, C. H. *et al. Phys. Rev. Lett.* **70**, 1895–1899 (1993).
- Ekert, A. K. *Phys. Rev. Lett.* **67**, 661–663 (1991).
- Horodecki, M., Oppenheim, J. & Winter, A. *Nature* **436**, 673–676 (2005).
- Hayden, P. *Nature* **436**, 633–634 (2005).

IMMUNOLOGY

In command of commensals

Humans must maintain a balanced composition for the trillions of commensal microbes that inhabit their gut, but how they do this is largely unclear. It now emerges that one factor is a molecular pathway in gut epithelial cells.

MENNO VAN LOOKEREN CAMPAGNE & VISHVA M. DIXIT

Homeostatic equilibrium between microbes in the gut lumen of the human host is crucial. Gut microbiota can aid digestion and suppress pathogenic bacteria. So when the spectrum of these resident microbes is altered, it can allow pathogenic species to elicit an inflammatory cascade in the gut mucosa, and inflammatory bowel disease can develop¹. Reporting in *Cell*, Elinav *et al.*² show that the protein NLRP6 in gut epithelial cells guards against harmful imbalances in the microbiota: in mice lacking this protein, microbial communities are altered, triggering both spontaneous and induced inflammation of the large intestine (colitis).

NLRP6 is related to three other proteins

— NLRP1, NLRP3 and NLRC4 — that respond to specific stimuli by forming large multi-protein complexes called inflammasomes. As part of the inflammasome, the enzyme caspase 1 triggers activation of the pro-inflammatory (inflammation-inducing) cytokine proteins IL-1 β and IL-18, which are then secreted³. An adapter protein known as ASC bridges the interaction between the NLR proteins and caspase 1 (Fig. 1a).

In a series of elegant experiments, Elinav and colleagues² show that genetic ablation of NLRP6 in mice profoundly reduces inflammasome activation in the intestine, alters the microbiota there and increases the animals' susceptibility to colitis. This severe intestinal inflammation results in significant weight loss. Moreover, the colitis-causing (colitogenic) microbiota could be transferred across

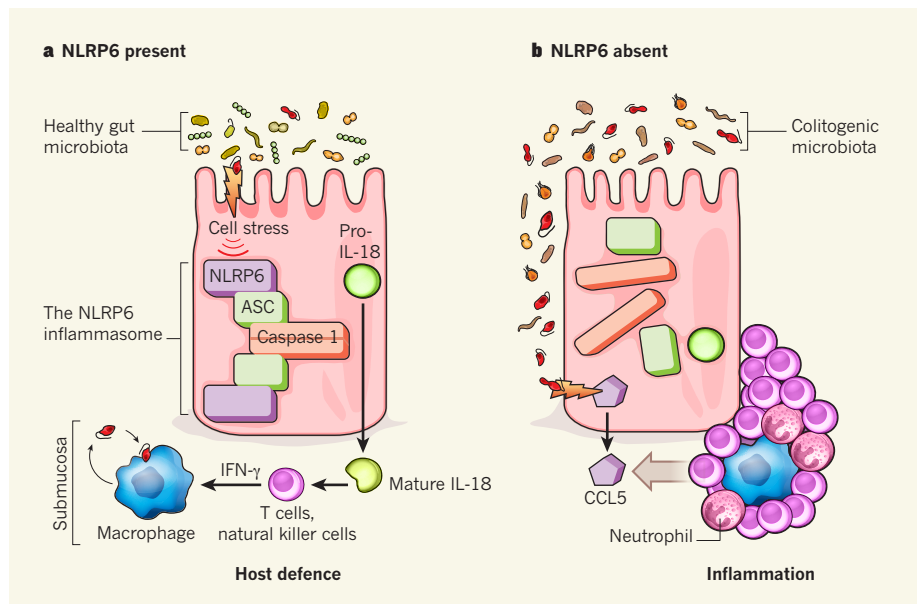


Figure 1 | The NLRP6 inflammasome — a guardian of intestinal homeostasis. **a**, Normally, intestinal epithelial cells respond to pathogenic bacteria that reside among commensal microbiota by mobilizing the NLRP6 inflammasome. Specifically, dimerization, and so activation, of caspase 1 results in the proteolytic conversion of pro-IL-18 into mature IL-18. The latter is secreted, stimulating production of IFN- γ in subsets of activated T cells or in natural killer cells in the submucosa. IFN- γ in turn promotes the bactericidal activity of macrophages. **b**, Elinav *et al.*² show that NLRP6-deficient epithelial cells fail to mount an appropriate response to pathogenic bacteria. Consequently, the normal microbial composition in the intestinal lumen is altered to that of a colitogenic community. These microbes stimulate epithelial cells to secrete the chemokine CCL5, which attracts immune cells such as neutrophils to trigger a chronic inflammatory response that can manifest as inflammatory bowel disease.

generations and could even displace the gut microbiota of co-housed mice that have normal immune systems. Colitogenic microbiota also developed in mice lacking ASC or caspase 1. These results underscore the central role of inflammasome activation downstream of NLRP6 in regulating the bacterial composition of the gut.

The authors provide genetic, morphological and electron-microscopic evidence that two bacterial species (from TM7 and Prevotellaceae) are over-represented in inflammasome-deficient mice. Colonies of these species invaded intestinal crypts — glands found in the epithelial lining of the small and large intestine that are generally devoid of pathogenic microbes. Elinav *et al.* further show that, in response to the altered microbiota, the chemokine protein CCL5 promotes infiltration of the large intestine by pro-inflammatory immune cells (Fig. 1b). Mice lacking CCL5 were less susceptible to colitis following exposure to faecal microbiota from NLRP6-deficient mice.

What are the implications of this work for inflammatory bowel disease (IBD) in humans? IBD develops through the interaction of at least three factors: environmental triggers, genetic predisposition and a deregulated immune response. Intestinal microbiota profoundly affect the host's immune status and are probably the most notable environmental factor in IBD⁴. Indeed, in the main types of IBD — Crohn's disease and ulcerative colitis — microbial imbalance is associated with a chronic inflammatory response in some patients¹.

Factors that might enhance a susceptibility to altered microbiota include genetic polymorphisms (variations) in proteins implicated in innate or adaptive immunity, stress within the cellular organelle known as the endoplasmic reticulum, and autophagy (the cell's degradation of its own components)¹. Of interest, the autophagy pathway acts upstream of inflammasome activation⁵. Polymorphisms in proteins involved in autophagy, including IRGM and ATG16L1, may therefore affect epithelial inflammasomes, albeit indirectly. These and other genetic pathways that determine microbiota composition can now be further explored using emerging genetic tools — including next-generation high-throughput sequencing of the microbial gene repertoire — to verify their contribution to the development of IBD.

Another question is how inflammasome activation and the release of IL-18 alter the composition of the intestinal microbiota. IL-1 β and IL-18 are proposed to be the main downstream effectors of the inflammasome³. Whereas IL-1 β is expressed mainly in immune cells called monocytes and macrophages, IL-18 is expressed in intestinal epithelial cells⁶. Following its caspase-1-mediated processing and secretion, IL-18 acts synergistically with the pro-inflammatory cytokines IL-12 and IL-15 to elicit production of the cytokine IFN- γ

by activated T helper cells and natural killer cells of the immune system⁷. IFN- γ is a potent inducer of the microbicidal activity of macrophages, and so may influence the make-up of the microbiota (Fig. 1).

Clearly, IL-18 is not the only effector of the NLRP6 inflammasome, because mice lacking NLRP6 or ASC develop a microbiota that is substantially different from that of mice lacking IL-18 (ref. 2). Thus, other pathways that link inflammasome activation to microbial composition in the intestinal lumen warrant investigation.

The inflammasomes that have been investigated show great selectivity for specific stimuli. These stimuli are either pathogen-associated molecules or host-derived factors released following cell or tissue damage³, and they are recognized by cytoplasmic sensors upstream of inflammasome activation. The signal that triggers activation of the NLRP6 inflammasome in epithelial cells is unknown, but may also involve molecules released after cell or tissue damage. Notably, Elinav *et al.*² find that deletion of inflammasome sensors other than NLRP6 (AIM2, NLRC4, NLRP10 and NLRP12) does not result in transferable colitogenic microbiota, and that NLRP3-deficient mice show attenuated colitis. Therefore, the nature of the initial stimulus that activates

the inflammasome is important for intestinal homeostasis.

Elinav and co-workers² present the NLRP6 inflammasome as a prime example of a host pathway that controls disease progression by driving specific alterations in the microbiota. Their results add another level of complexity to the development of IBD and deepen our understanding of the interplay between host, microbiota and genetic susceptibility factors. Such knowledge should pave the way for improved therapies for this debilitating range of disorders. ■

Menno van Lookeren Campagne and Vishva M. Dixit are at Genentech, South San Francisco, California 94080, USA.
e-mails: menno@gene.com;
dixit.vishva@gene.com

1. Kaser, A., Zeissig, S. & Blumberg, R. S. *Annu. Rev. Immunol.* **28**, 573–621 (2010).
2. Elinav, E. *et al. Cell* doi:10.1016/j.cell.2011.04.022 (2011).
3. Schroder, K. & Tschopp, J. *Cell* **140**, 821–832 (2010).
4. Mayer, L. J. *Gastroenterol.* **45**, 9–16 (2010).
5. Nakahira, K. *et al. Nature Immunol.* **12**, 222–230 (2011).
6. Zaki, M. H. *et al. Immunity* **32**, 379–391 (2010).
7. Dinarello, C. A. *Annu. Rev. Immunol.* **27**, 519–550 (2009).

The authors declare competing financial interests.
See go.nature.com/dppqxa

PALAEOANTHROPOLOGY

In search of the australopithecines

Evidence from strontium isotope ratios preserved in fossil teeth provides a glimpse into the group dynamics and ranging habits of the australopithecines that can be compared with the patterns for modern primates. SEE LETTER p.76

MARGARET J. SCHOENINGER

The australopithecines, human relatives who lived during the Pliocene and early Pleistocene epochs roughly 3.5 million to 1.8 million years ago, remain enigmatic creatures. The trouble is that inferences about their biology and lifestyle have to be drawn from frustratingly little hard information, which means that conjectures about them are plentiful but our knowledge remains scant.

On page 76 of this issue, Copeland and colleagues¹ describe the results of their investigation into a new aspect of australopithecine life. They analysed strontium isotope data from the tooth enamel of two australopithecine species, *Australopithecus africanus* and *Paranthropus robustus*, recovered from fossils in two caves in South Africa. The ratio of two isotopes of strontium (⁸⁷Sr/⁸⁶Sr) depends on the type of

soils in the environment and is unaltered by biological processes. Australopithecine food items record the same strontium ratio as the soil in which they grew. Significantly, the ratio varies between types of bedrocks around the two cave sites, Sterkfontein and Swartkrans (see Fig. 1 of the paper¹).

On the basis of the ⁸⁷Sr/⁸⁶Sr ratios for 19 individuals, Copeland *et al.*¹ conclude that, in both species, most of the larger individuals (presumably males) fed primarily near the cave sites, whereas several of the smaller individuals (presumably female) fed elsewhere. The authors suggest that this indicates male philopatry (males remain in their natal group), with female dispersal. The situation in living African apes provides one point of comparative reference: chimpanzees have male philopatry, whereas both sexes of mountain gorillas may disperse².

The isotope ratios imply that males limited

their home and day ranges to a circumscribed area. The $^{87}\text{Sr}/^{86}\text{Sr}$ values seen in the male teeth derive from a dolomite-rich region that is only 30 square kilometres in extent, distributed in a northeast–southwest direction. The minimum distances from the caves to bedrock of differing ratios are only 2–3 km to the southeast and 5–6 km to the northwest. There are no barriers, yet the males stayed in the dolomitic areas, apparently because their foods existed only in such regions. Young females apparently fed in other, unspecified areas during tooth mineralization, and moved into the males' region as young adults for reproductive purposes.

These results have implications for understanding australopithecine diet, group size, predator avoidance and home-range size. Living primate species offer insights into these variables². Among modern primate groups, ecology influences the temporal and geographic patterns of exploitation of food items, group size and predator pressure, and all of these factors strongly influence group ranging patterns. For example, leaf-eating primates have smaller home ranges than fruit-eating primates of the same body size. Body size and group size influence when and where individual primates would be vulnerable to predators.

Australopithecines lived in woodland savannahs, which provide tree foods and grass³. The carbon stable isotope ratios ($^{13}\text{C}/^{12}\text{C}$) in tooth enamel suggest that most of them ate both foodstuffs, with some East African members eating only grass^{4,5}. Their tooth-enamel structure and its microscopic scratches and pits imply that they also ate hard objects (seeds and nuts) or plant tubers and swollen stems⁶.

This diet contrasts with that of extant chimpanzees, which select only tree foods when living in savannah regions⁷. In such habitats, chimpanzees elude predators by avoiding open areas, forming larger than normal groups, and covering large home ranges³. For example, the range of chimpanzees at Ugalla, Tanzania³, is some 600 km² in extent, much larger than the 30 km² proposed by Copeland *et al.*¹ for the South African australopithecines. The small home ranges suggested by the $^{87}\text{Sr}/^{86}\text{Sr}$ ratios argue against australopithecines having targeted only tree foods and in favour of their including a more ubiquitous food such as grass. Mountain gorillas⁸ feed on ubiquitous terrestrial vegetation in a home range of up to 25 km², similar to that suggested by the strontium isotope data for male australopithecines¹. But male gorillas are very large, thereby providing group protection, and they live in forest habitats without large predators². In contrast to the mountain gorilla, the australopithecines lived in an open habitat, had a smaller body size and faced many species of large carnivores. How did they avoid predation while eating grass in an open savannah?

For diet, the closest equivalent to that of the australopithecines is found in the extant baboon *Theropithecus gelada* and in its giant fossil ancestor⁹, which had a body size

roughly equivalent to that of the australopithecines⁵ and which occupied the same or a similar habitat. Modern *T. gelada* are much smaller than either fossil group and practise male philopatry; they avoid predation in their open savannah habitat by living in bands up to 60 strong, and have overlapping home ranges of sizes roughly equivalent to that estimated for the australopithecines (see references in ref. 2). But how could two large primate species of similar body size, one quadrupedal (giant baboon) and the other bipedal (australopithecine), have coexisted? They would have been competing with each other, as well as with other fruit-eating animals, and with grass-eaters such as pigs, hippos and zebra ancestors, along with other ungulate species⁵. How the australopithecines balanced predator avoidance and the need to compete for food remains an open question.

In Lewis Carroll's flight of nonsensical whimsy, *The Hunting of the Snark*¹⁰, there is the warning about the day that "your Snark be a Boojum!" — for the consequence of that is instant disappearance. We are in the midst of a fascinating hunt, but for us a Snark that turns out to be a Boojum would mean success, not

failure. What is needed in this field of research is for conjectures about the australopithecines to vanish, instead of arising time and time again^{5,9}. We must hope that continued investigation along the lines of Copeland and colleagues' study¹ will one day allow us to winnow away some of that speculative thinking. ■

Margaret J. Schoeninger is in the Department of Anthropology and the Center for the Advanced Research and Training in Anthropogeny, University of California at San Diego, La Jolla, California 92093, USA. e-mail: mjschoen@ucsd.edu.

1. Copeland, S. R. *et al.* *Nature* **474**, 76–78 (2011).
2. Strier, K. B. *Primate Behavioral Ecology* 3rd edn (Allyn & Bacon, 2007).
3. Moore, J. J. in *Topics in Primatology* Vol. 1 (eds Nishida, T. *et al.*) 99–118 (Univ. Tokyo Press, 1992).
4. Lee-Thorp, J. & Sponheimer, M. *Am. J. Phys. Anthropol.* **131**, Suppl. 43 Yb. *Phys. Anthropol.* 131–148 (2006).
5. Cerling, T. *et al.* *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.1104627108 (2011).
6. Lucas, P. *et al.* *BioEssays* **30**, 374–385 (2008).
7. Schoeninger, M. J., Moore, J. & Sept, J. M. *Am. J. Primatol.* **49**, 297–314 (1999).
8. Watts, D. P. *Int. J. Primatol.* **19**, 651–680 (1998).
9. Jolly, C. J. *Man* **5**, 5–26 (1970).
10. Carroll, L. *The Hunting of the Snark: An Agony in Eight Fits* (Macmillan, 1898).

MOLECULAR EVOLUTION

Hidden diversity sparks adaptation

Although some genetic mutations have clear effects, others have been considered neutral and inconsequential. Such cryptic mutations can nonetheless facilitate adaptation to new environments. SEE LETTER P.92

JEREMY A. DRAGHI & JOSHUA B. PLOTKIN

Is genetic variation primarily neutral, or is it predominantly shaped by natural selection? Biologists have debated this question for decades. And although the neutralist–selectionist argument has fuelled advances in our understanding of molecular evolution, it has contributed less to our knowledge of how organisms adapt to changing environments. Recent theories about robustness and evolvability, as well as experiments with enzymes and with microbes, suggest that adaptation may depend on a deep and largely unappreciated interaction between neutral and beneficial mutations. On page 92 of this issue, Hayden *et al.*¹ describe an empirical test of this hypothesis.

When exposed to a new environment, a microbial population will rapidly adapt through the emergence and fixation of beneficial mutations². Nevertheless, microbial evolution experiments have uncovered factors

that complicate this simple picture. One such factor, called epistasis — whereby the effects of one mutation are modified by other mutations — explains why some populations adapt faster or more predictably than others. These differences in evolvability, or in the propensity to produce beneficial mutations, account for several surprising results (such as the sudden emergence of a major metabolic adaptation³ or the long-term coexistence of lineages in a population⁴). The renewed interest in epistasis also points to an exciting and constructive role for neutral mutations in the process of adaptation.

The key finding that links neutral mutations to beneficial changes is that neutrality is often conditional: a mutation may have no detectable effect when it arises in a given environment and against a specific genetic background, but subsequent changes to the environment or genome may reveal hidden fitness effects⁵. Theoretical work suggests that conditionally neutral, or 'cryptic', mutations may accumulate within a population and later be expressed

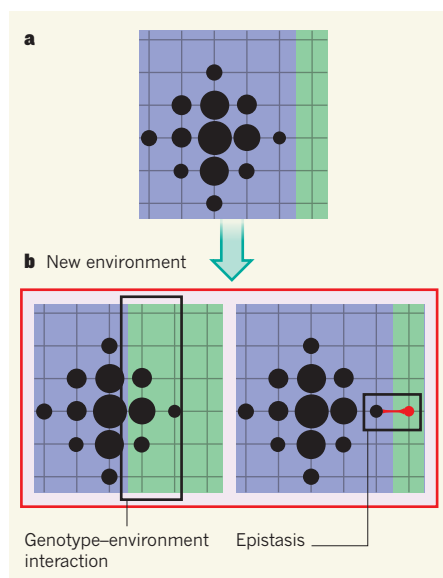


Figure 1 | Role of cryptic genetic variation in adaptation. In this depiction, genotypes (nodes) are separated by mutations (lines). Node size indicates the frequency of a genotype in the population, and background colour shows the phenotype in a given environment. **a**, A population under selection for the blue phenotype occupies a network of equally fit genotypes. **b**, When the population is challenged with a new selective environment (red box), some of the previously neutral genotypes may already express the adaptive phenotype (green). Adaptation to the new environment is also facilitated by epistasis between a beneficial mutation (red) and a mutation that was neutral in the original environment. Hayden *et al.*¹ confirm that neutral genetic variation accumulated in one environment can facilitate adaptation to a novel environment, either through interaction between genotype and the environment or by epistasis.

after an environmental change, either directly in response to the new environment⁶, or by epistatic interactions with subsequent mutations⁷ (Fig. 1). Hayden *et al.*¹ test this theory by direct experimentation.

The authors evolved two populations of a ribozyme (RNA enzyme) under selection for activity with its native substrate; one population was subject to strong selection, the other to weaker selection. The populations accumulated cryptic genetic variation that was later revealed when they were challenged to catalyse a different substrate. By measuring the increase in catalytic activity over time in this new selective environment, Hayden and colleagues found that the rate of adaptation was proportional to the amount of cryptic variation present in each population. In other words, cryptic genetic variation that had accumulated in one environment facilitated adaptation to a new environment.

This *in vitro* ribozyme system facilitates precise measurements of genetic variation throughout experiments. It also allows the mutations responsible for adaptation to be identified, sequenced and reconstructed in the

ancestral ribozyme strain. To interrogate the genetic basis of adaptation, Hayden *et al.* took advantage of these features. They confirmed that the beneficial genotypes arose from previously cryptic genetic variation, supporting a causal connection between neutral mutations and adaptation.

The analysis¹ further revealed a complex, ecological dynamic: in one of the populations, the proper functioning of an evolved genotype depended on the presence of other genotypes in the population. So by tracking population variation through deep sequencing, the researchers were able to capture a more detailed picture of the underlying genetics and ecology than is feasible in most evolution experiments using organisms with large genomes.

Hayden *et al.*¹ applied a very high per-base mutation rate to their short ribozyme sequence to approximate the typical genomic mutation rate of a living organism. But extrapolating their results to real organisms is not straightforward, because there might be different patterns of epistasis among the sites of a single enzyme from those among the genes in a large genome. Nevertheless, other work^{8–10} suggests that epistasis is prevalent at the genomic scale, and that novel environments reveal previously cryptic phenotypic variation in complex organisms. Together with Hayden and colleagues' work, these studies of natural populations suggest a general and important role for cryptic variation in determining a population's adaptive potential.

Sewall Wright, a vigorous proponent of the importance of epistasis in evolution, presaged the modern view in his comment¹¹ on Kimura's neutral theory: "Changes in wholly nonfunctional parts of the molecule would be the most frequent ones but would be unimportant, unless they occasionally give a basis for later changes which improve function in the

species in question which would then become established by selection." Indeed, Hayden *et al.*¹ have now verified that cryptic variation can shape the process of adaptation. We can hope that future studies of this type will elucidate the specific mechanisms by which neutral mutations potentiate adaptation (see ref. 12, for instance), and so eventually lead to a quantitative understanding of a population's rate of adaptation.

As climate change brings organisms into new environments and facilitates contact between novel pathogens and humans, a comprehensive theory for how populations adapt will have pay-offs for conservation and public health, as well as for our understanding of the diversity of life. ■

Jeremy A. Draghi and Joshua B. Plotkin
are in the Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

e-mail: jplotkin@sas.upenn.edu

- Hayden, E. J., Ferrada, E. & Wagner, A. *Nature* **474**, 92–95 (2011).
- Elena, S. F. & Lenski, R. E. *Nature Rev. Genet.* **4**, 457–469 (2003).
- Blount, Z. D., Borland, C. Z. & Lenski, R. E. *Proc. Natl Acad. Sci. USA* **105**, 7899–7906 (2008).
- Woods, R. J. *et al.* *Science* **331**, 1433–1436 (2011).
- Wagner, A. *Robustness and Evolvability in Living Systems* (Princeton Univ. Press, 2005).
- Hermisson, J. & Wagner, G. P. *Genetics* **168**, 2271–2284 (2004).
- Draghi, J. A., Parsons, T. L., Wagner, G. P. & Plotkin, J. B. *Nature* **463**, 353–355 (2010).
- Masel, J. & Trotter, M. V. *Trends Genet.* **26**, 406–414 (2010).
- Schlichting, C. D. *Ann. NY Acad. Sci.* **1133**, 187–203 (2008).
- Jarosz, D. F. & Lindquist, S. *Science* **330**, 1820–1824 (2010).
- Provine, W. B. *Sewall Wright and Evolutionary Biology* (Chicago Univ. Press, 1986).
- Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. *Proc. Natl Acad. Sci. USA* **103**, 5869–5874 (2006).

CLIMATE CHANGE

Ancient Antarctic fjords

The East Antarctic ice sheet, the largest in the world, lies seemingly frozen in time. Discovery of a rugged landscape buried beneath the thick ice provides evidence of a more dynamic past. SEE LETTER P.72

SANDRA PASSCHIER

Ice sheets reflect 90% of incoming solar radiation and are of considerable mass, and hence affect Earth's global geophysical parameters. To gain insight into the mechanisms of the potentially long response times of ice sheets, it is necessary to study them from a long-term geological perspective. In Antarctica, reconstructing ice-sheet drainage patterns at times of past ice-sheet minima is a challenge:

today 98% of the continent is buried under several kilometres of ice. Airborne geophysical surveys that penetrate the ice have helped to meet that challenge, and have contributed significantly to what is known about Antarctica's bed topography. In reconstructing the subglacial topography of the deep Aurora Subglacial Basin in East Antarctica, Young *et al.*¹ (page 72 of this issue) have made a notable discovery by revealing subglacial landscapes in astonishing detail.

From the monitoring of Antarctic ice sheets using monthly satellite gravity data, it seems that the sheets are losing mass, and that this mass loss has accelerated during the past decade². Coupled ice-sheet and climate modelling of long-term changes suggest that, in contrast to the West Antarctic ice sheet, the East Antarctic ice sheet is generally not susceptible to melting beyond its marine-based margins³. However, uncertainty remains about the effect of the topography of the Wilkes and Aurora subglacial basins⁴, where the East Antarctic ice sheet lies on a landward-dipping bed in water that is more than 1 kilometre deep in places. In this configuration, ice sheets are susceptible to unstable retreat, and the resulting ice loss could contribute to accelerated sea-level rise.

Using ice-penetrating radar data collected by the ICECAP aerogeophysical programme, Young *et al.*¹ surveyed the preserved topography under thick ice in the Aurora Subglacial Basin. Their studies reveal a previously unknown subglacial mountain range dissected by a series of parallel valleys; these valleys have U-shaped cross-sections and are some 50 kilometres wide. Rock uplands form barriers to ice flow (Fig. 1), and knowing their location sets an important boundary condition for ice-sheet models. The newly discovered subglacial mountains, at one time positioned at the edge of the ice sheet, would have restricted the outflow of ice, and outlet glaciers with steep gradients carved deep into the rock. A fjord landscape resulted — perhaps similar to that seen in East Greenland or Norway today — in which glacial valleys were carved over several glacial periods and, at times, were flooded by the sea.

Young *et al.*¹ provide the first detailed geophysical evidence of upland erosion by outlet glaciers of the East Antarctic ice sheet at a time when the ice edge was more than 400 kilometres inland of its present location. The upland erosion must have taken place when the East Antarctic ice sheet was dynamic and had oscillated to a size much smaller than that documented for the current interglacial stage (the past 12,000 years or so). The magnitude of glacial erosion would have required the ice to carry significant subglacial meltwater for at least part of each year. The low erosive capacity of the currently slow-moving ice, and the fact that the ice-flow direction is oblique to the orientation of the glacial valleys, rules out the possibility that the fjords identified by Young *et al.* resulted from erosion under the ice sheet that covers the mountains today.

When did the East Antarctic ice sheet last display dynamic behaviour? Young *et al.*¹ favour an age older than 14 million years for the carving of the glacial valleys, although no new independent age control is presented. The authors cite compelling evidence of atmospheric cooling at high altitude in the Dry Valleys of the Transantarctic Mountains⁵ to support a hypothesis involving ice-sheet stability, a lack of meltwater and persistent frigid



S. PASSCHIER/NSF

Figure 1 | Ancient Antarctic topography. This landscape in the Royal Society Range of the Transantarctic Mountains was sculpted long ago by rivers and glaciers. A polar alpine glacier with limited erosive power now occupies the valley. The Transantarctic Mountains form a barrier to the outflow of ice from the East Antarctic ice sheet. Young *et al.*¹ have discovered that a similar mountain barrier is preserved beneath several kilometres of ice in the Aurora Subglacial Basin.

conditions in East Antarctica since that time.

However, this hypothesis must be reconciled with other evidence^{6,7} showing elevated sea-surface temperatures in the nearby Southern Ocean, and the possibility of surface melt across portions of East Antarctica with some ice retreat in the Aurora Subglacial Basin⁴ for a period of global warmth between about 3 million and 5 million years ago. Unfortunately, well-dated sediment archive⁸ and proxy evidence⁹ for East Antarctic ice-sheet dynamism is biased towards periods before about 16 million years ago and is simply lacking for more recent times. So, for the moment, the age of the newly discovered glacial topography and the implied dynamic ice sheet remains uncertain.

The main point of Young and colleagues' paper, however, is that the East Antarctic ice sheet once had a dynamic, oscillating ice edge within one of the enigmatic subglacial basins. The paper increases the level of detail at which we can resolve the bed topography for the Aurora Subglacial Basin and will help in making improvements to ice-sheet models. These models must be validated with independent data sets, such as well-dated sediment archives, which are scarce in and around East Antarctica. Data gathered by the recently completed

Expedition 318 of the Integrated Ocean Drilling Program¹⁰, which has been drilling offshore of the nearby Wilkes Subglacial Basin, should start to remedy this deficiency. ■

Sandra Passchier is in the Department of Earth and Environmental Studies, Montclair State University, Montclair, New Jersey 07043, USA.
e-mail: passchiers@mail.montclair.edu

1. Young, D. A. *et al.* *Nature* **474**, 72–75 (2011).
2. Velicogna, I. *Geophys. Res. Lett.* doi:10.1029/2009GL040222 (2009).
3. Pollard, D. & DeConto, R. M. *Nature* **458**, 329–332 (2009).
4. Hill, D. J., Haywood, A. M., Hindmarsh, R. C. A. & Valdes, P. J. in *Deep-Time Perspectives on Climate Change* (eds Williams, M., Haywood, A. M., Gregory, J. & Schmidt, D. N.) 517–538 (Geol. Soc. Lond., 2007).
5. Lewis, A. R., Marchant, D. R., Ashworth, A. C., Hemming, S. R. & Machlus, M. L. *GSA Bull.* **119**, 1449–1461 (2007).
6. Escutia, C. *et al.* *Glob. Planet. Change* **69**, 170–184 (2009).
7. Whitehead, J. M. & Bohaty, S. M. *Paleoceanography* doi:10.1029/2002PA000829 (2003).
8. Naish, T. R. *et al.* *Nature* **413**, 719–722 (2001).
9. Pekar, S. F. & DeConto, R. M. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **231**, 101–109 (2006).
10. Expedition 318 Scientists *IODP Prelim. Rep.* doi:10.2204/iodp.pr.318.2010 (2010).

Crystal structure of the FimD usher bound to its cognate FimC–FimH substrate

Gilles Phan^{1*}, Han Remaut^{1,2*}, Tao Wang^{3*}, William J. Allen^{1*}, Katharina F. Pirker¹, Andrey Lebedev⁴, Nadine S. Henderson⁵, Sebastian Geibel¹, Ender Volkan⁶, Jun Yan¹, Micha B. A. Kunze¹, Jerome S. Pinkner⁶, Bradley Ford^{6,7}, Christopher W. M. Kay^{1,8,9}, Huilin Li^{3,10}, Scott J. Hultgren⁶, David G. Thanassi⁵ & Gabriel Waksman^{1,9}

Type 1 pili are the archetypal representative of a widespread class of adhesive multisubunit fibres in Gram-negative bacteria. During pilus assembly, subunits dock as chaperone-bound complexes to an usher, which catalyses their polymerization and mediates pilus translocation across the outer membrane. Here we report the crystal structure of the full-length FimD usher bound to the FimC–FimH chaperone–adhesin complex and that of the unbound form of the FimD translocation domain. The FimD–FimC–FimH structure shows FimH inserted inside the FimD 24-stranded β -barrel translocation channel. FimC–FimH is held in place through interactions with the two carboxy-terminal periplasmic domains of FimD, a binding mode confirmed in solution by electron paramagnetic resonance spectroscopy. To accommodate FimH, the usher plug domain is displaced from the barrel lumen to the periplasm, concomitant with a marked conformational change in the β -barrel. The amino-terminal domain of FimD is observed in an ideal position to catalyse incorporation of a newly recruited chaperone–subunit complex. The FimD–FimC–FimH structure provides unique insights into the pilus subunit incorporation cycle, and captures the first view of a protein transporter in the act of secreting its cognate substrate.

Gram-negative pathogens commonly interact with their environment using long, linear, surface-exposed protein appendages called pili. In uropathogenic *Escherichia coli*, type 1 pili carry at their distal end a dedicated mannose-specific adhesin, FimH, that is responsible for the attachment of bacteria to the bladder epithelium and their subsequent internalization and biofilm-like organization inside the urothelial cells.

Type 1 pili are representative of a large class of non-covalently linked fibres on the surface of gram-negative bacteria, synthesized via the conserved chaperone/usher pathway^{1–3}. Type 1 pili are composed of four different subunit types (FimH, FimG, FimF and FimA). The adhesin FimH and two linker subunits FimG and FimF form a short flexible fibrillar tip that is attached to an extended rigid and helically wound rod of thousands of FimA subunits (Supplementary Fig. 1a)^{4–6}. Subunits cross the inner membrane via the SecYEG secretory pathway. In the periplasm, folding and stability of the subunits require formation of a binary complex with the FimC chaperone^{7,8}. Chaperone–subunit complexes are then targeted to the outer membrane usher, FimD, which catalyses the ordered polymerization of subunits and enables the translocation of the growing fibre across the outer membrane in a self-energized process^{9,10}.

All pilus subunits (or pilins) exhibit an incomplete Ig-like fold, characterized by the absence of the C-terminal β -strand^{11–13} (Supplementary Fig. 1b), leaving a deep hydrophobic groove on the subunit surface (Supplementary Fig. 1c). As a result, pilus subunits are unstable on their own, unless in a chaperone–subunit complex or bound to an adjacent subunit within the pilus. Both chaperone–subunit and subunit–subunit interactions involve a fold-complementation mechanism whereby the subunit's non-canonical Ig-fold is complemented

in trans by, respectively, an extended β -strand in the N-terminal domain of the chaperone (strand G1) or a 10 to 20-residue-long peptide extension at the N terminus of the adjacent subunit (called the N-terminal extension or Nte)^{11–14} (Supplementary Fig. 1b). During subunit polymerization, the chaperone donor strand binding the subunit's hydrophobic groove (an interaction termed donor-strand complementation or DSC) is replaced by the Nte of the newly incorporated subunit in a process called donor-strand exchange (DSE)¹¹ (Supplementary Fig. 1b).

The structure of the translocation domain of the P pilus usher PapC in its inactive state revealed a 24-stranded β -barrel protein¹⁵. The loop between strands 6 and 7 of the β -barrel holds an 80-residue insertion that forms a plug domain that, in the non-engaged usher, resides in the barrel lumen, gating the usher channel shut. In addition to the translocation domain, ushers (~800 residues) contain a ~120-residue N-terminal domain (NTD) responsible for chaperone–subunit binding and recruitment^{16–18} and a ~170 residue C-terminal domain (CTD) of poorly understood function^{19,20} (Fig. 1a). How these domains cooperate to recruit chaperone–subunit complexes, catalyse subunit polymerization, and translocate the nascent pilus through the membrane is unknown. To provide insights into these processes, we present here the crystal structure of the FimD usher bound to its cognate FimC–FimH chaperone–adhesin substrate and that of the non-engaged FimD usher translocation domain.

Structure of the FimD–FimC–FimH complex

A stoichiometric complex containing the type 1 pilus usher FimD bound to the FimC–FimH chaperone–adhesin complex (Fig. 1a)

¹Institute of Structural and Molecular Biology, University College London and Birkbeck College, Malet Street, London WC1E 7HX, UK. ²Structural & Molecular Microbiology and Structural Biology Brussels, VIB - Vrije Universiteit Brussels, 1050 Brussels, Belgium. ³Biology Department, Brookhaven National Laboratory, Upton, New York 11973, USA. ⁴Department of Chemistry, University of York, York YO10 5YW, UK. ⁵Center for Infectious Diseases and Department of Molecular Genetics & Microbiology, Stony Brook University, Stony Brook, New York 11794, USA. ⁶Department of Molecular Microbiology and Center for Women's Infectious Disease Research, Washington University School of Medicine, Saint Louis, Missouri 63110, USA. ⁷Department of Pathology and Immunology, Washington University, St Louis, Missouri 63110, USA. ⁸London Centre for Nanotechnology, University College London, 17–19 Gordon Street, London WC1H 0AH, UK. ⁹Research Department of Structural and Molecular Biology, University College London, Gower Street, London WC1E 6BT, UK. ¹⁰Department of Biochemistry and Cell Biology, Stony Brook University, Stony Brook, New York 11794, USA.

*These authors contributed equally to this work.

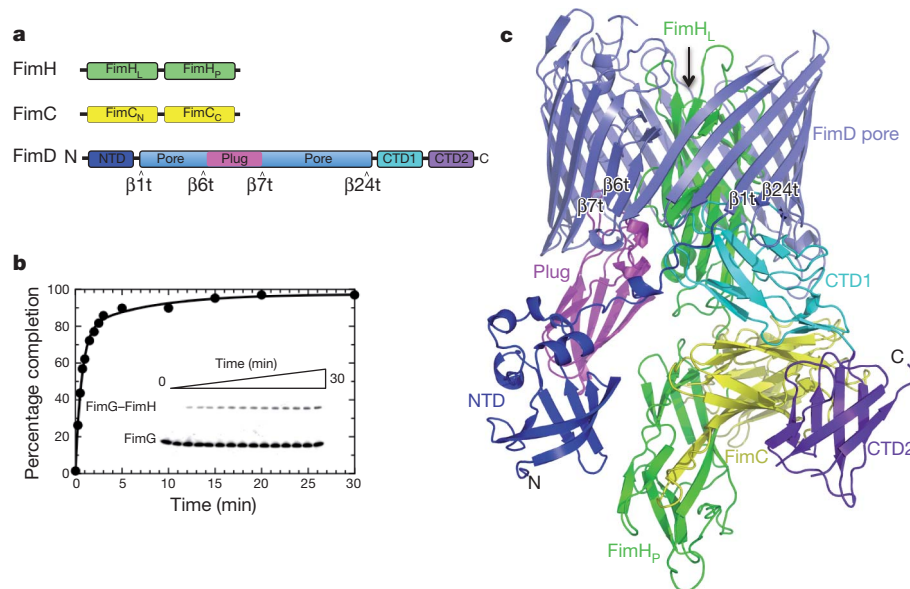


Figure 1 | Structure of the FimD-FimC-FimH complex. **a**, Schematic diagram of domain organization of FimH (FimH_L, lectin domain; FimH_P, pilin domain) and FimC (FimC_N, N-terminal domain; FimC_C, C-terminal domain) and FimD (see text). **b**, Activity assay demonstrating that the purified FimD-FimC-FimH complex is functional. FimD-FimC-FimH was challenged at $t = 0$ by the FimC-FimG^{S92C[A647]} complex fluorescently labelled by Alexa 647 reacted on residue 92 of FimG (see position of residue 92 in Supplementary Fig. 2b). Intensity of the fluorescent FimG-FimH band (the DSE product) was used to

was purified and shown to be active (Fig. 1b). It was then crystallized and its structure determined to 2.8 Å resolution (Fig. 1c, Supplementary Fig. 2a, Supplementary Table 1 and Methods). Like PapC, FimD contains a 24-stranded β-barrel (residues 139–665), interrupted by a plug domain (residues 241–324) inserted in the periplasmic loop linking strands 6 and 7 (Figs 1, 2, and topology diagram in Supplementary Fig. 3). However, in contrast to the PapC structure, which captured the non-activated, unbound translocation channel, the plug domain in the FimD-FimC-FimH complex now resides in the periplasm, underneath the translocation domain and next to the NTD (Fig. 1c and Supplementary Fig. 4). The usher NTD has been shown to form a binding site for chaperone-subunit complexes, including FimC-FimH^{16–18}. In the FimD-FimC-FimH structure, however, the NTD lies idle, making no interactions with FimC (see below); the FimC-FimH complex instead is bound to two Ig-like domains formed at the usher C terminus, CTD1 and CTD2 (residues 666–750 and 751–834, respectively).

FimH is a two-domain protein (Fig. 1a), where the N-terminal lectin domain (residues 1–157; FimH_L) is responsible for receptor binding, and the C-terminal or pilin domain (residues 158–279; FimH_P) forms the interacting region with either the chaperone within the chaperone-adhesin complex in the periplasm or with the adjacent subunit (FimG) within the pilus¹². In the ternary FimD-FimC-FimH complex, FimC stabilizes the FimH pilin domain via a typical DSC fold-complementation interaction, which remains unchanged compared to the FimC-FimH complex alone¹². Remarkably, the FimH lectin domain inserts into the lumen of the translocation channel, traversing the entire length of the channel, its tip exposed on the extracellular side of the usher. FimD is the first transporter to be visualized with a substrate protein inserted through its lumen. The FimH pilin domain and the FimC chaperone are located underneath the pore.

Usher activation involves a large conformational change in the β-barrel domain

The FimC-FimH complex is the first chaperone-subunit complex to bind to the usher and is required to drive a conformational change in

assess the percentage progress of the DSE reaction. Inset, raw SDS-PAGE gel visualized as described in Methods. Each band represents a time point. **c**, Side view ribbon representation of the FimD-FimC-FimH structure, with FimH in green, FimC in yellow and the FimD NTD, β-barrel, plug, CTD1 and CTD2 in blue, slate, magenta, cyan and purple, respectively. β1t, β6t and β7t, and β24t indicate the β-barrel strands (see secondary structure labelling nomenclature in Supplementary Fig. 3a) connecting the barrel to the NTD, the plug and the CTDs, respectively.

the latter that primes it for pilus biogenesis^{10,21,22}. The molecular nature of this activation process is unknown. To get a direct comparison between the FimC-FimH-engaged form and the apo form of the type 1 pilus usher, we crystallized the isolated FimD translocation domain (residues 124–663) and determined its structure to 3.0 Å resolution

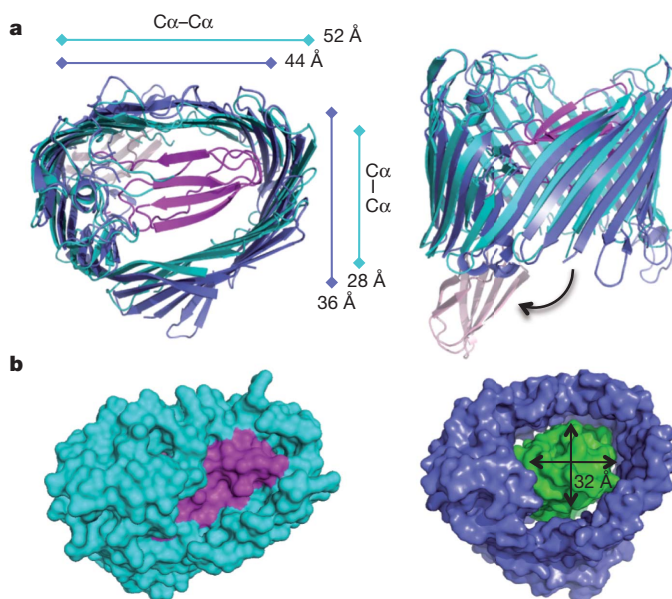


Figure 2 | Channel conformations in apo and activated (FimC-FimH-engaged) FimD usher. **a**, Top (left) and side (right) view ribbon representations of the superimposed apo-FimD (cyan) and activated FimD (slate) β-barrel. The plug domain in the channel lumen in apo FimD (magenta) rotates into the periplasm following FimD activation (pink). **b**, Top view surface representation of the apo-FimD (left) and activated FimD (right, for clarity, showing only the translocation channel and FimH lectin domain, FimH_L). The plug and FimH_L are coloured magenta and green, respectively.

(Fig. 2 and Supplementary Fig. 5a). Apo-FimD closely resembles the structure of the PapC translocation domain (r.m.s.d. (root mean squared deviation) for corresponding C α atoms of 1.7 Å). It is composed of a kidney-shaped 24-stranded β -barrel occluded by a plug domain (residues 241–324). The structure of the translocation domain in the FimC–FimH-engaged usher shows a marked conformational change in the β -barrel. The 24-stranded β -barrel rearranges from an oval-shaped pore with a 52 Å by 28 Å diameter to a near circular pore of 44 Å by 36 Å diameter (C α to C α distances; Fig. 2a, left panel). This large conformational rearrangement in the FimD translocation channel upon activation by FimC–FimH is unprecedented in β -barrel proteins, which were until now considered rigid structures.

In the apo-FimD, the translocation channel is completely sealed off by the plug domain (Fig. 2b, left panel). In the FimC–FimH-engaged complex, the plug domain is displaced into the periplasm, opening a circular channel of 32 Å now occupied by the FimH lectin domain (Fig. 2b, right panel). In apo-FimD the plug domain makes close contacts with the inner wall of the β -barrel, burying 2,738 Å² of surface area (Fig. 2b, left panel). In contrast, in the ternary complex, the β -barrel–FimH interface buries 1,590 Å² of surface area and includes fewer contacts with FimH compared to the β -barrel–plug interface in the apo form (Fig. 2b, right panel): only 6 β -barrel C α atoms lie within 5 Å from FimH in FimD–FimC–FimH, compared to 39 β -barrel C α atoms lying within 5 Å of the plug in apo-FimD. The more distant contact in the ternary complex structure probably provides room for the variability in subunit diameter among the different subunit types and also might facilitate translocation through the pore (Supplementary Fig. 5c, d).

Usher contains two chaperone–subunit binding sites

So far, the only region of the usher known to bind chaperone–subunit complexes is the usher N-terminal domain (NTD)^{16–18}. The FimD–FimC–FimH structure now shows the existence of a second binding site on the usher, located at the C-terminal domains, CTD1 and CTD2 (Figs 1c and 3a). The FimC–FimH complex contacts the FimD usher over a surface area of 3,802 Å². Apart from the interaction of the FimD channel with the FimH lectin domain (see above), the most extensive interaction with the FimC–FimH complex is formed by the usher CTD1 (Fig. 3a and Supplementary Fig. 6a). CTD1 contacts the FimH lectin domain and FimC over a surface area of 621 Å² and 422 Å², respectively. Contact area between CTD2 and the FimC–FimH complex is 504 Å² large and is primarily with FimC. Removal of the CTDs or of CTD2 alone or point mutations in CTD1 abrogate pilus biogenesis (see ref. 23 and this work (Supplementary Table 2)). Using electron paramagnetic resonance (EPR) spectroscopy we also demonstrate that subsequent subunits localize to the CTDs binding site after undergoing DSE (Fig. 3c and Supplementary Fig. 7). Moreover, these complexes are fully functional, that is, able to incorporate the next subunit into the nascent pilus (Supplementary Figs 2b, c).

Other than its interaction with the CTDs, the FimC–FimH complex also comes into contact with the usher plug domain and the NTD (Fig. 3b and Supplementary Fig. 6b). The contact surface area between the plug and the FimH lectin domain is significant (474 Å²). Although the NTD is located within proximity of the FimH pilin domain, the small contact surface area of 189 Å² and its low shape complementarity²⁴ of 0.45 indicate a weak interaction (Supplementary Fig. 6b). Notably, this contact zone does not overlap with the known, canonical chaperone–subunit binding site at the NTD (see later and Supplementary Fig. 6c).

When comparing the interface between FimD CTDs and FimC–FimH in the FimD–FimC–FimH structure with the interface between the FimD NTD and FimC–FimH in the structure of the NTD–FimC–FimH_p complex reported previously¹⁷, it becomes apparent that the binding sites overlap (Supplementary Fig. 8).

Thus, the usher contains two chaperone–subunit binding sites and the question arises whether these are mutually exclusive for chaperone–subunit binding or rather operate in concert, and if so, in what sequence.

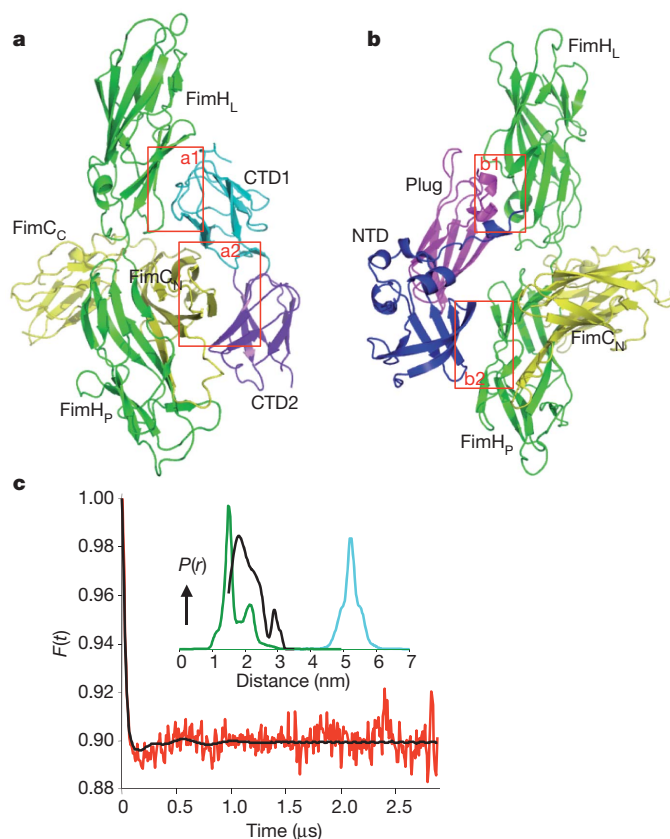


Figure 3 | FimC–FimH interactions with FimD in the FimD–FimC–FimH complex. **a, b,** Side view of the ribbon representation of the FimC–FimH interface with the FimD CTDs (**a**) and with the FimD plug and NTD, as found in the FimD–FimC–FimH complex (**b**). For clarity, only the respective FimD domains are shown. The boxed interfaces (a1, FimH–CTD1; a2, FimC–CTD2; b1, FimH–plug and b2, FimH–NTD) are described in the text and shown in detail in Supplementary Fig. 6. Colour coding is as in Fig. 1. **c,** DEER measurement of the distance between two nitroxide spin labels, one on residue 756 of FimD (located in CTD2) in the FimD–FimC–FimH complex, and the other on residue 74 of FimC in the FimC–FimG complex (see details and controls in Methods and Supplementary Fig. 7; see also Supplementary Fig. 7 for results of distance measurements by EPR between residue 774 of FimD CTD2 and residue 74 of FimC). The Form factor (main graph; red line), the fit to the data using DeerAnalysis2010 (main graph; black line; ref. 38), and the distance distribution derived from the data (inset; black line) are shown. For comparison, we include the distance distribution predicted by MMM³⁹ from the crystal structure of FimD–FimC–FimH, assuming that the position of FimC–FimG is similar to the previously bound chaperone–subunit complex FimC–FimH (green line) and the distance distribution from a model structure of FimD–FimC–FimH where FimC–FimG was positioned at the NTD as in ref. 17 (cyan line; see Supplementary Fig. 7a). It can be seen that the vast majority of the distance distribution obtained experimentally overlaps with that predicted when FimC–FimG locates at the CTDs. A minor fraction corresponding to a distance around 3 nm suggests a conformational equilibrium in solution.

A single usher protomer forms a pilus assembly machine

Chaperone/usher pili extend by step-wise addition of new chaperone–subunit complexes at the base of the growing fibre. Because the last incorporated chaperone–subunit complex is known to remain bound on the usher^{18,22}, the usher requires two chaperone–subunit binding sites for function. The FimD–FimC–FimH structure and the EPR data presented here demonstrate that subunits at the base of the fibre are bound to the CTDs, with the NTD lying idle. To investigate whether in the FimD–FimC–FimH complex the NTD is able to recruit the next chaperone–subunit complex, we superimposed the known structure of the FimD NTD bound to FimC–FimF²⁵ (the structure of the NTD–FimC–FimG complex is not available) onto the NTD in the FimD–FimC–FimH crystal structure (Fig. 4a, b). This superimposition

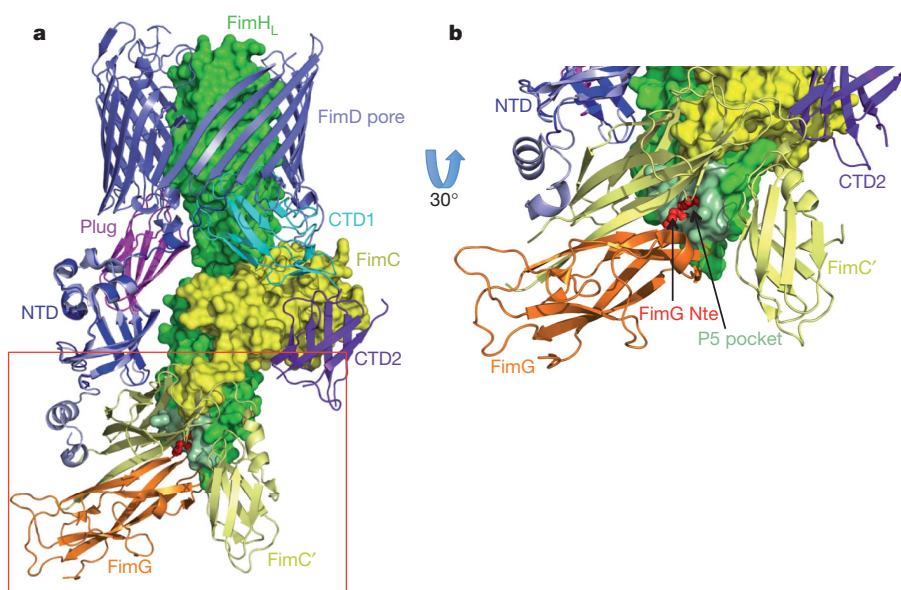


Figure 4 | Chaperone-subunit incorporation cycle at the FimD usher. **a**, Side view of the FimD-FimC-FimH complex (FimC-FimH in surface representation) with a new chaperone-subunit complex (FimC'-FimG, yellow-orange, respectively) modelled at the NTD binding site (the model is from PDB 3BWU; that is, based on the crystal structure of FimD NTD alone

demonstrates that the NTD in the FimD-FimC-FimH complex is available for recruitment of a chaperone-subunit complex without steric clashes with the FimC-FimH complex bound at the CTDs. The requirement of an accessible NTD was tested by an *in vitro* DSE experiment, where the chaperone-subunit binding site of the NTD of the purified FimD-FimC-FimH complex was blocked by a bulky molecule (Supplementary Fig. 9). The inactivation of the NTD results in a near loss of further subunit incorporation, indicating that the NTD indeed acts as the recruitment site for new chaperone-subunit complexes^{16–18}.

The superimposition presented in Fig. 4a provides unique insights into the catalytic mechanism of a monomeric usher. The ability of the Nte of an incoming subunit to initiate the DSE reaction with the previously assembled subunit is crucially dependent on a defined binding site in that subunit, called the P5 site²⁶ (Supplementary Fig. 1). The P5 site allows the incoming Nte to access the hydrophobic groove of the preceding subunit, allowing it to displace the chaperone donor strand in a step-wise zip-in-zip-out mechanism^{26–29}. The *in silico* model of FimC-FimF docked at the NTD of the FimD-FimC-FimH complex shows that the newly recruited subunit comes into close proximity with the FimH pilin domain, representative for the subunit that resides at the base of the growing fibre (Fig. 4a). Strikingly, the Nte of the subunit bound at the NTD lies directly above the P5 pocket of the subunit bound at the CTDs, perfectly positioned to initiate the DSE reaction (Fig. 4b). Together, the active recruitment of new chaperone-subunit complexes to the usher NTD and their ideal positioning with respect to the penultimate chaperone-subunit complex located at the CTDs provide a rationale for the catalytic ability of the usher (Supplementary Fig. 10). In the proposed model for the catalytic cycle, the chaperone-subunit complex at the base of the growing pilus fibre resides at the usher's CTDs. New subunits are recruited to the NTD and brought into ideal orientation to undergo DSE with the subunit bound at the CTDs (now the penultimate subunit; Supplementary Fig. 10, step 1). Upon DSE, the chaperone is displaced from the penultimate subunit and dissociates from the CTDs (Supplementary Fig. 10, step 2). To reset the assembly machinery for a new incorporation, the incoming chaperone-subunit complex would need to dissociate from the NTD and be transferred to the CTDs site, concomitantly pushing the penultimate subunit into the translocation channel (Supplementary Fig. 10,

bound to FimC-FimF). **b**, Clipped view of the (FimC'-FimG)-(FimC-FimH) contact zone (boxed area in **a**), showing positioning of the FimG N-terminal extension (FimG Nte; in red) above the P5 pocket in the FimC-FimH complex (FimC-FimH in yellow-green, the P5 pocket shown in light green).

steps 3 and 4, respectively). How the hand-over of the chaperone-subunit complex from the usher's NTD to the CTDs occurs remains speculative.

Conclusion

The crystal structure of FimD bound to FimC-FimH provides the remarkable view of a protein transporter caught in the act of secreting its cognate substrate. Together with the FimD translocator domain structure, it elucidates not only the mechanism of gating leading to FimH insertion into the FimD barrel, but also the subsequent steps of subunit polymerization and nascent pilus translocation. Pilicide compounds recently shown to inhibit pilus biogenesis target the interface between chaperone-subunit complexes and the usher NTD³⁰. The crystal structure presented here unravels a complex choreography of domain motion and protein-protein interactions that will no doubt be of crucial importance in the design of additional compounds capable of disrupting type 1 pilus biogenesis and thus inhibiting cystitis, an infectious disease that plagues millions of individuals worldwide.

METHODS SUMMARY

Purification and crystallization. FimD-FimC-FimH with a Strep-tag at the C terminus of FimD was purified as described previously with an additional Strep-tag affinity chromatography step¹⁵. After addition of trypsin (which removes 21 residues at the N terminus of FimD and cleaves its β 13–14 loop), the complex was crystallized by hanging-drop vapour diffusion. The 6×His-tagged FimD translocation domain (residues 124–663) was purified by Ni-NTA affinity and size exclusion chromatography, and crystallized by hanging-drop vapour diffusion.

Structure determination and refinement. The crystals of the FimD-FimC-FimH complex contained two ternary complexes per asymmetric unit, related by a pseudotranslation. The chaperone-subunit (FimC-FimH) or usher domains, for which the structures (NTD) or structures of homologous domains (PapC translocation domain and plug, PapC CTD2) were available, were located individually using molecular replacement, as implemented in Phaser³¹ and Molrep³². CTD1 was built manually using Coot³³. Refinement with Refmac^{34,35} converged to a model with an *R* factor of 0.219 and an *R*_{free} of 0.277. The structure of the FimD translocation domain was solved by molecular replacement with the equivalent PapC structure (PDB code 2VQI) as a search model using Phaser³¹. The structure was built in Coot³³, and refined in Phenix³⁶ to an *R* factor of 0.229 and *R*_{free} of 0.305. **DSE assay.** The FimD-FimC-FimH complex was mixed with fluorescently labelled FimC-FimG, where FimG was labelled with Alexa 647 on FimG residue 92. DSE progression was monitored by the appearance of the fluorescent FimG-FimH band

on SDS–PAGE gels. For DSE experiments involving a FimD–FimC–FimH complex with a bulky molecule blocking NTD binding, FimD was reacted with Alexa 594 on residue 109.

EPR spectroscopy. The FimD–FimC–FimH complex was spin-labelled on residue 756 or residue 774 of FimD. The FimC–FimG complex was spin-labelled on residue 74 of FimC. Double electron–electron resonance (DEER) measurements for distance determination were performed as described previously³⁷.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 September 2010; accepted 13 April 2011.

- Mulvey, M. A. *et al.* Induction and evasion of host defenses by type 1-piliated uropathogenic *Escherichia coli*. *Science* **282**, 1494–1497 (1998).
- Sauer, F. G., Remaut, H., Hultgren, S. J. & Waksman, G. Fiber assembly by the chaperone–usher pathway. *Biochim. Biophys. Acta* **1694**, 259–267 (2004).
- Waksman, G. & Hultgren, S. J. Structural biology of the chaperone–usher pathway of pilus biogenesis. *Nature Rev. Microbiol.* **7**, 765–774 (2009).
- Jones, C. H. *et al.* FimH adhesin of type 1 pili is assembled into a fibrillar tip structure in the Enterobacteriaceae. *Proc. Natl Acad. Sci. USA* **92**, 2081–2085 (1995).
- Hahn, E. *et al.* Exploring the 3D molecular architecture of *Escherichia coli* type 1 pili. *J. Mol. Biol.* **323**, 845–857 (2002).
- Le Trong, I. *et al.* Structural basis for mechanical force regulation of the adhesin FimH via finger trap-like β sheet twisting. *Cell* **141**, 645–655 (2010).
- Vetsch, M. *et al.* Pilus chaperones represent a new type of protein-folding catalyst. *Nature* **431**, 329–333 (2004).
- Barnhart, M. M. *et al.* PapD-like chaperones provide the missing information for folding of pilin proteins. *Proc. Natl Acad. Sci. USA* **97**, 7709–7714 (2000).
- Jacob-Dubuisson, F., Striker, R. & Hultgren, S. J. Chaperone-assisted self-assembly of pili independent of cellular energy. *J. Biol. Chem.* **269**, 12447–12455 (1994).
- Nishiyama, M., Ishikawa, T., Rechsteiner, H. & Glockshuber, R. Reconstitution of pilus assembly reveals a bacterial outer membrane catalyst. *Science* **320**, 376–379 (2008).
- Sauer, F. G. *et al.* Structural basis of chaperone function and pilus biogenesis. *Science* **285**, 1058–1061 (1999).
- Choudhury, D. *et al.* X-ray structure of the FimC–FimH chaperone–adhesin complex from uropathogenic *Escherichia coli*. *Science* **285**, 1061–1066 (1999).
- Zavialov, A. V. *et al.* Structure and biogenesis of the capsular F1 antigen from *Yersinia pestis*: preserved folding energy drives fiber formation. *Cell* **113**, 587–596 (2003).
- Sauer, F. G., Pinkner, J. S., Waksman, G. & Hultgren, S. J. Chaperone priming of pilus subunits facilitates a topological transition that drives fiber formation. *Cell* **111**, 543–551 (2002).
- Remaut, H. *et al.* Fiber formation across the bacterial outer membrane by the chaperone/usher pathway. *Cell* **133**, 640–652 (2008).
- Nishiyama, M., Vetsch, M., Puorger, C., Jelesarov, I. & Glockshuber, R. Identification and characterization of the chaperone–subunit complex-binding domain from the type 1 pilus assembly platform FimD. *J. Mol. Biol.* **330**, 513–525 (2003).
- Nishiyama, M. *et al.* Structural basis of chaperone–subunit complex recognition by the type 1 pilus assembly platform FimD. *EMBO J.* **24**, 2075–2086 (2005).
- Ng, T. W., Akman, L., Osisami, M. & Thanassi, D. G. The usher N terminus is the initial targeting site for chaperone–subunit complexes and participates in subsequent pilus biogenesis events. *J. Bacteriol.* **186**, 5321–5331 (2004).
- Shu Kin So, S. & Thanassi, D. G. Analysis of the requirements for pilus biogenesis at the outer membrane usher and the function of the usher C-terminus. *Mol. Microbiol.* **60**, 364–375 (2006).
- Ford, B. *et al.* Structural homology between the C-terminal domain of the PapC usher and its plug. *J. Bacteriol.* **192**, 1824–1831 (2010).
- Munera, D., Hultgren, S. & Fernandez, L. A. Recognition of the N-terminal lectin domain of FimH adhesin by the usher FimD is required for type 1 pilus biogenesis. *Mol. Microbiol.* **64**, 333–346 (2007).
- Saulino, E. T., Thanassi, D. G., Pinkner, J. S. & Hultgren, S. J. Ramifications of kinetic partitioning on usher-mediated pilus biogenesis. *EMBO J.* **17**, 2177–2185 (1998).
- Li, Q. *et al.* The differential affinity of the usher for chaperone–subunit complexes is required for assembly of complete pili. *Mol. Microbiol.* **76**, 159–172 (2010).
- Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
- Eidam, O., Dworkowski, F. S., Glockshuber, R., Grutter, M. G. & Capitani, G. Crystal structure of the ternary FimC–FimF₁–FimD_N complex indicates conserved pilus chaperone–subunit complex recognition by the usher FimD. *FEBS Lett.* **582**, 651–655 (2008).
- Remaut, H. *et al.* Donor-strand exchange in chaperone-assisted pilus assembly proceeds through a concerted β strand displacement mechanism. *Mol. Cell* **22**, 831–842 (2006).
- Verger, D., Miller, E., Remaut, H., Waksman, G. & Hultgren, S. Molecular mechanism of P pilus termination in uropathogenic *Escherichia coli*. *EMBO Rep.* **7**, 1228–1232 (2006).
- Rose, R. J. *et al.* Unraveling the molecular basis of subunit specificity in P pilus assembly by mass spectrometry. *Proc. Natl Acad. Sci. USA* **105**, 12873–12878 (2008).
- Verger, D. *et al.* Structural determinants of polymerization reactivity of the P pilus adaptor subunit PapF. *Structure* **16**, 1724–1731 (2008).
- Pinkner, J. S. *et al.* Rationally designed small compounds inhibit pilus biogenesis in uropathogenic bacteria. *Proc. Natl Acad. Sci. USA* **103**, 17897–17902 (2006).
- McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D* **63**, 32–41 (2007).
- Vagin, A. & Teplyaev, A. Molecular replacement with MOLREP. *Acta Crystallogr. D* **66**, 22–25 (2010).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
- Collaborative Computational Project, number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Pannier, M., Veit, S., Godt, A., Jeschke, G. & Spiess, H. W. Dead-time free measurement of dipole–dipole interactions between electron spins. *J. Magn. Reson.* **142**, 331–340 (2000).
- Jeschke, G. *et al.* DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. *Appl. Magn. Reson.* **30**, 473–498 (2006).
- Polyhach, Y., Bordignon, E. & Jeschke, G. Rotamer libraries of spin labelled cysteines for protein studies. *Phys. Chem. Chem. Phys.* **13**, 2356–2366 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgments This work was funded by Medical Research Council grant 85602 to G.W., NIH grant GM62987 to D.G.T., NIH grants 49950, 29549 and 48689 to S.J.H., and NIH grant GM74985 and BNL LDRD grant 10-16 to H.L.; H.R. is supported by a VIB Young PI project grant and the Odysseus program of the FWO-Vlaanderen. K.F.P. is supported by a Schrödinger Fellowship from the Austrian Science Fund, project J 2959-N17. We thank the staff of beamlines X25 and X29 at NSLS, the staff of beamline ID23-1 at ESRF, N. Cronin for technical assistance during data collection, and H. Saibil and E. Orlova for comments on the manuscript.

Author contribution G.P. produced the FimD–FimC–FimH complex, grew the crystals of this complex, collected X-ray crystallographic data, and initiated the determination of the structure by molecular replacement, and participated in the building and refinement of the structure. H.R. produced the FimD–FimC–FimH complex, trained G.P., supervised the work, analysed the structures and wrote the paper. T.W. grew crystals of the FimD translocation domain, collected X-ray crystallographic data, and determined the structure. W.J.A. set up the DSE assay and prepared the samples for EPR. K.F.P. carried out the EPR experiments, which were analysed by K.F.P., M.B.A.K. and C.W.M.K.; A.L. completed the structure determination of the FimD–FimC–FimH complex, built and refined the structure. N.S.H., E.V., J.S.P. and B.F. cloned and purified the translocation domain of FimD, and cloned and analysed the FimD CTD mutants. S.G. participated in the building and refinement of the FimD–FimC–FimH structure and analysed the structure. J.Y. carried out the native mass spectrometry experiments on the FimD–FimC–FimH complex. C.W.M.K. supervised the EPR work. H.L., S.J.H. and D.G.T. supervised the work on apo-FimD, analysed the structures, and wrote the paper. G.W. supervised the work on FimD–FimC–FimH, analysed the structures, and wrote the paper.

Author Information Structure factors and coordinates have been deposited in the Protein Data Bank (entry codes 3RFZ and 30HN for coordinates and structure factors of the FimD–FimC–FimH complex and the translocation domain of FimD, respectively). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.W. (g.waksman@ucl.ac.uk or g.waksman@bbk.ac.uk) or D.G.T. (david.thanassi@stonybrook.edu).

METHODS

Expression and purification of the outer membrane FimD–FimC–FimH complex. *Escherichia coli* strain B834 (Novagen) was transformed with two plasmids: pETS1001 encoding *fimC_{His}* under arabinose control and pAN2 encoding *fimD* under IPTG control²². A *strep-tag* II (SA-WSHPQFEK) was added to the C terminus of FimD by the SLIM protocol (site-directed ligase independent mutagenesis⁴⁰; Supplementary Table 3). Bacteria were grown in TB media containing kanamycin and spectinomycin at 37 °C. At $D_{600} = 1.0$, the culture was induced with 100 μ M isopropyl- β -D-thiogalactoside (IPTG) and 0.1% (w/v) L-arabinose with a supplement of 0.1% (v/v) glycerol. The induced bacteria were grown for 48 h at 16 °C.

Outer membranes were prepared as described in ref. 15. Outer membrane proteins were solubilized in 20 mM Tris-HCl, pH 8.5, 120 mM NaCl, 1.5% (w/v) dodecylmaltopyranoside (DDM; Anatrace) and protease inhibitors cocktail (Calbiochem) for 30 min at room temperature. The extract was cleared by ultracentrifugation (45 min at 100,000g, 4 °C), loaded onto a streptavidin column, washed with 100 mM Tris-HCl, pH 8.5, 120 mM NaCl, 0.05% (w/v) DDM, and the bound fraction eluted with the same buffer containing 2.5 mM D-desthiobiotin.

Limited proteolysis of the purified FimD–FimC–FimH complex was carried out by adding directly trypsin (Sigma) to the *strep-tag* II affinity eluted fraction, with a ratio of 1:50 (w/w) of enzyme to substrate for 3 h at room temperature. Trypsin removes 21 amino acids at the N terminus of the FimD usher (cut after R21) and cleaves the usher translocation domain at loop β 13–14 after residue K469, as assessed by N-terminal sequencing. Overall, the trypsin-digested complex has a molecular mass of 141 kDa compared to 144 kDa for the undigested complex (both molecular masses were assessed by mass spectrometry). Such a very minor trimming of the complex was crucial to obtain crystals, presumably removing sequences preventing crystal packing. The digested FimD–FimC–FimH was loaded onto a nickel affinity column, washed with 20 mM Tris-HCl, pH 8.5, 120 mM NaCl, 0.05% (w/v) DDM and 25 mM imidazole, detergent-exchanged with 20 mM Tris-HCl, pH 8.5, 120 mM NaCl, 2 mM LDAO (Anatrace) and 25 mM imidazole, and eluted with that same buffer containing 250 mM imidazole. 0.8% (v/v) of tetraethylene glycol monooctyl ether (C8E4; Anatrace) was then added to the nickel-affinity-eluted fraction before concentration using a 100 kDa cut-off spin concentrator (Amicon) and loading onto a HiLoad Sephacryl S300 16/60 (GE Healthcare) gel filtration column in 20 mM Tris-HCl, pH 8.5, 50 mM NaCl, 2 mM LDAO and 0.8% (v/v) C8E4. The digested FimD–FimC–FimH complex eluted as a single peak and was concentrated using a 100 kDa cut-off spin concentrator (Amicon).

Expression and purification of the FimD translocation domain. The FimD translocation domain (residues 124–663) was identified by mass spectroscopy of the limited trypsin treatment product of purified full-length FimD, and constructed using the SLIM method⁴⁰ from parental plasmid pETS4 (ref. 22), which encodes *fimD*-6 \times His under IPTG control (Supplementary Table 3). The final plasmid, pNH297, encodes the native FimD signal sequence followed by the translocation domain followed by a short linker sequence (GGPVAT), thrombin cleavage site (LVPRGS) and 6 \times His-tag.

After induction, outer membranes were obtained as described in ref. 15. Proteins were extracted from the outer membranes with 1.5% (w/v) DDM (Anatrace) in a buffer containing 25 mM Tris, pH 8.2, 300 mM NaCl, 10% (v/v) glycerol, and 1 \times protease inhibitors cocktail (Roche) at 4 °C overnight. The mixture was ultracentrifuged (100,000g, 60 min, 4 °C) to remove debris. Supernatant was loaded onto a 5-ml Ni-NTA cartridge (Qiagen) pre-equilibrated in 25 mM Tris, pH 8.2, 300 mM NaCl, 0.05% (w/v) DDM, and 20 mM imidazole. Detergent exchange was performed at this step by washing the column with 25 mM Tris, pH 8.2, 300 mM NaCl, 0.8% (w/v) C8E4 (Anatrace). The target protein was eluted in the same buffer containing a step gradient of imidazole (20 mM, 50 mM, and 300 mM). After further purification by size exclusion chromatography (Superdex-200, GE Healthcare), the FimD translocation domain was concentrated to 10–15 mg ml⁻¹ in 5 mM Tris, pH 8.2, 50 mM NaCl, 1.5% (w/v) C8E4 (Anatrace).

Crystallization and data collection of the FimD–FimC–FimH complex. Trypsin-digested FimD–FimC–FimH complex crystals were grown using the vapour diffusion method at 20 °C. The crystallization drops contained 6–9 mg ml⁻¹ of purified complex ($D_{280} = 8$ –13), 50 mM ammonium acetate, pH 6.0–8.5, 4% (v/v) isopropanol and 770–840 mM ammonium sulphate. After 18 days, needle or blade-like crystals were flash-cooled in liquid nitrogen using the mother liquor with 30% (v/v) glycerol as cryoprotectant.

The data were collected at ESRF beamline ID23-1 (Grenoble, France) and were processed to 2.8 Å resolution using MOSFLM⁴¹. The integrated data were merged using POINTLESS and SCALA⁴². Space group, cell dimensions, and data collection statistics are reported in Supplementary Table 1. There was a strong non-origin

peak in the Patterson map with the height of 0.37 relative to the height of the origin peak. This peak corresponded to the pseudo-translation $1/2c \pm \delta b$ with $\delta \approx 2.5$ Å.

Crystallization and data collection of the FimD translocation domain. The FimD translocation domain was crystallized by hanging-drop vapour diffusion method at 21 °C. Protein solution were mixed by 1:1 ratio with well solution. Plate-like crystals appeared under condition of 100 mM Na citrate, pH 4.8–6.5, 7% (w/v) PEG 4000, 100 mM NaCl, 50 mM MgCl₂, 20 mM spermine HCl. Crystals were flash-cooled in liquid nitrogen using mother liquor containing 30% (v/v) MPD as cryoprotectant.

Data were collected at beamline X25 at the National Synchrotron Light Source and processed to 3.0 Å resolution with HKL2000 (ref. 43). Space group, cell dimensions and data collection statistics are reported in Supplementary Table 1.

Structure determination and refinement of the FimD–FimC–FimH complex. The structure contains nine types of different domains in three different polypeptide chains (FimC and FimH contain two domains each and FimD contains five domains). Structural information was available for all individual domains but one, CTD1. The method used for location of the first three structural units (FimC, FimH, and the translocation domain of FimD) was the standard molecular replacement search (equivalent to the search in the Patterson map) implemented in both Phaser³¹ and Molrep³². The three methods used for location of the plug, NTD and CTD2 were variants of the search in the electron density map implemented in Molrep. All three methods use $2F_o - F_c$ type maps from a refined partial structure, map coefficients from Refmac³⁴ being used in this work. (1) The first method uses conventional rotation function (RF) against structure amplitudes from the map masked by the partial structure to find orientation of the model, and the phased translation function (PTF) to find its position. (2) The second method uses spherically averaged phased translation function (SAPTF⁴⁴) to generate a list of possible positions of the centre of mass of the model, phased rotation function (PRF) to assign an orientation to each potential position and PTF to verify and correct the position of the model. (3) The third method differs from the second one in that the PRF is replaced by the standard rotation function against structure amplitudes from the electron density in a sphere around the tested position of the centre of mass. In addition, for each of the three methods, the positions of two pseudotranslation-related copies of a model were being searched for simultaneously or one after another, and cross-checked using the clear translational peak in the native Patterson. There were no homologues with known structure for CTD1 of FimD and this domain was built manually using Coot³³ when all other structural units were located.

To locate the FimC–FimH complex, PDB codes 1KLF and 3BWU were used. Two copies were found using the standard molecular replacement. The next unit to be located was the translocation/barrel domain of FimD. The equivalent PapC domain (PDB code 2VQI) was positioned with both Phaser and Molrep using the previously found FimC–FimH substructure as a fixed model.

The resulting model did not refine well, probably because of conformational differences between bound and unbound structures. Fortunately, the latest version of Refmac³⁴ offered a “jelly body” refinement, which in contrast to conventional refinement favoured locally correlated changes in the atomic parameters. The “jelly body” refinement was applied to the partial structure containing FimC, FimH and the translocation domain of FimD and substantially changed the shape of the barrel, C- α atoms being shifted up to 3.8 Å.

The plug domain search model was from PDB entry code 2VQI. The six modes of Molrep described above were tried. The solution found with methods (2) and (3) placed the boundary residues of the plug in close proximity to the translocation domain residues to which the plug domain must be connected. Refinement resulted in a sensible electron density map leaving little doubts that the solution was correct.

One copy of the NTD of FimD (the search model derived from PDB entry code 1EZ3) was found by both Phaser and Molrep in all six Molrep’s modes tried. The second copy was found with five out of six Molrep modes. In contrast, the CTD2 of FimD (the search model from 3I48, sequence identity 32%), which had poorer electron density than all other domains even in the final structure, was only located using method (2) including SAPTF, PRF and PTF. Moreover, one of the two CTD2 domains is less ordered than the other and the location of this domain required simultaneous search for two pseudotranslation-related copies.

Model building of the FimD–FimC–FimH complex was carried out manually in Coot³³. Restrained refinement where no σ cutoff was applied was performed in Refmac 5.6 (ref. 34), including different NCS group restraints for each protein domains related by the pseudo-translation. The following regions had poor density and thus are not part of the final FimD model: F22–G25, S188–K195 (loop β 3–4), G454–Y473 (loop β 13–14) and E805–N807. A small loop in the final FimC model is also missing: S179–G182. At the end, 95% of the FimD–FimC–FimH model was built. Refinement statistics of the final model are reported in Supplementary Table 1.

Structure determination and refinement of the FimD translocation domain. Molecular replacement was carried out using the Phaser-Phenix³⁶ program and the PapC monomer (PDB entry code 2VQI) as search model. The FimD usher translocation domain was manually rebuilt in Coot³³ and refined (no σ cutoff applied) in Phenix. An N-terminal fragment (124–138), a middle loop (454–471), a C-terminal fragment (657–663), and the linker plus the 6×His tag were disordered and could not be traced in the model. The refinement statistics are listed in Supplementary Table 1.

DSE assay. A single cysteine mutation was introduced at position 92 of FimG (termed hereafter FimG^{S92C}) using the Quikchange site-directed mutagenesis protocol (Stratagene; Supplementary Table 3). The FimC–FimG^{S92C} complex was expressed and purified as described previously for wild-type FimC–FimG²² then labelled with Alexa 647–C₂-maleimide (Invitrogen). The labelling reaction was carried out by incubating 100 μ M protein and 160 μ M fluorophore together overnight at 4 °C in a buffer consisting of 50 mM Tris-HCl pH 8.0, 150 mM NaCl and 1 mM EDTA. Excess dye was removed by gel filtration (Superdex-75 column from GE Healthcare) in labelling buffer, yielding pure FimC–FimG^{S92C[A647]} as assessed by SDS–PAGE. Final protein concentration was determined using an extinction coefficient at 280 nm of 35,066 M^{−1} cm^{−1}, after correcting for the absorbance of Alexa 647 at 280 nm. Typical labelling efficiencies were between 80 and 100%.

Full-length FimD–FimC–FimH complex for donor strand exchange was purified as described above, with the following exceptions: (1) no detergent exchange was carried out on the nickel affinity column, instead the protein was eluted in a buffer consisting of 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.05% DDM and 500 mM imidazole; and (2) the final gel filtration step was performed in a buffer consisting of 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.05% DDM and 1 mM EDTA. The final concentration of FimD–FimC–FimH was determined using an extinction coefficient at 280 nm of 194,780 M^{−1} cm^{−1} (based on a 1:1:1 stoichiometry within the complex, an assumption confirmed by analytical ultracentrifugation (results not shown)).

To initiate donor strand exchange, 160 nM purified FimD–FimC–FimH was mixed rapidly with 1 μ M of FimC–FimG^{S92C[A647]} at 4 °C, in a buffer consisting of 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA and 0.05% DDM. Aliquots of reaction mix were quenched at various time intervals by mixing 10:1 with 2 M HCl. After adding SDS–PAGE loading buffer (but not boiling as boiling disrupts subunit–subunit interaction), the FimG^{S92C[A647]}–FimH product (identified by mass spectrometry) was separated from the FimG^{S92C[A647]} substrate by SDS–PAGE. Note that FimC–FimH alone, in the absence of usher, does not react with FimC–FimG^{S92C[A647]} within the time frame of the experiment. The fluorescent gel bands were visualized using an FLA-3000 fluorescence plate reader (Fujifilm), with excitation at 633 nm and a long-pass emission cutoff of 675 nm. Bands corresponding to FimG^{S92C[A647]} (FimG) and FimG^{S92C[A647]}–FimH (FimG–FimH) were selected and quantified using Image Gauge (Fujifilm), and the background fluorescence subtracted from each band. Product formation was calculated by the equation:

$$[\text{GH}] = [\text{G}_{\text{tot}}] \frac{I_{\text{GH}}}{I_{\text{GH}} + I_{\text{G}}}$$

where [GH] is the concentration of FimG^{S92C[A647]}–FimH product formed, I_{GH} and I_{G} are the corrected intensities of the FimG–FimH and FimG bands respectively, and $[\text{G}_{\text{tot}}]$ is the initial concentration of FimC–FimG^{S92C[A647]} used. [GH] data were converted to percentage completion by:

$$\text{Percentage completion} = 100 \frac{[\text{GH}]}{[\text{DCH}]}$$

where [DCH] is the initial concentration of FimD–FimC–FimH.

Blocking the NTD of FimD. To block the chaperone–subunit binding site on the NTD, a mutation to Cys was introduced at residue 109 of FimD (Supplementary Table 3). The purified FimD^{Q109C}–FimC–FimH complex was reacted with Alexa 594 maleimide for 1 h on ice. Alexa 594 is here used as a block. DSE assay was carried out as described above using wild-type FimD–FimC–FimH, Alexa 594-labelled wild-type FimD–FimC–FimH (to control for the effect of non-specific labelling), FimD^{Q109C}–FimC–FimH and Alexa 594-labelled FimD^{Q109C}–FimC–FimH.

Formation of the fluorescent FimG^{S92C[A647]}–FimH band was monitored as above.

EPR spectroscopy. EPR distance measurements were carried out to determine the position of the FimC–FimG chaperone–subunit complex relative to the usher CTD2 in solution. This was achieved by site-directed spin labelling of the FimD–FimC–FimH and FimC–FimG complexes with a nitroxide spin label (1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl methanethiosulphonate (MTSSL)). Cysteine residues were introduced at position 74 of FimC in the FimC–FimG complex, and separately at positions 756 and 774 of FimD in the FimD–FimC–FimH complex by QuikChange site-directed mutagenesis (Stratagene; Supplementary Table 3). FimC^{Q74C}–FimG was expressed and purified as described for the DSE assay, then labelled with MTSSL, using 20 μ M protein, 400 μ M MTSSL and the same buffer conditions as for fluorescent labelling (see above). FimD^{T756C}–FimC–FimH and FimD^{S774C}–FimC–FimH mutants did not have a strep-tag present; they were therefore expressed and purified as for the DSE assay but with the strep-tag affinity column omitted. Labelling was carried out before the final gel filtration step, using the same protocol as for FimC^{Q74C}–FimG but with the addition of 0.05% DDM to the labelling buffer. All mutants were exchanged into D₂O buffer to enhance the transverse relaxation time of the electron spins, which enables measurement of longer distances. The concentration of spin label was determined and corresponded to a labelling efficiency in the range 70–100%. The estimated error for the spin label efficiency is approximately $\pm 15\%$ due to errors in the determination of the protein concentration and the determination of the double integral of the EPR spectra.

Solutions of 70 μ M FimD^{T756C[MTSSL]}–FimC^{Q74C[MTSSL]}–FimG–FimH and 100 μ M FimD^{S774C[MTSSL]}–FimC^{Q74C[MTSSL]}–FimG–FimH (50 μ l) were prepared by mixing FimD^{T756C[MTSSL]}–FimC–FimH or FimD^{S774C[MTSSL]}–FimC–FimH with FimC^{Q74C[MTSSL]}–FimG in a ratio of 1:1. Glycerol (5%) was added as cryoprotectant. The mixture was transferred into a quartz capillary of 2 mm (inner diameter) and frozen in liquid nitrogen. Controls included mixing FimD^{T756C[MTSSL]}–FimC–FimH or FimD^{S774C[MTSSL]}–FimC–FimH with unlabelled FimC^{Q74C}–FimG or mixing unlabelled FimD^{T756C}–FimC–FimH or FimD^{S774C}–FimC–FimH with labelled FimC^{Q74C[MTSSL]}–FimG in a ratio of 1:1.

Continuous-wave EPR experiments were performed at 160 K on a Bruker EMXplus spectrometer operating at 9.4 GHz equipped with a 4122SHQE resonator and an Oxford Instruments ESR900 cryostat. All measurements were carried out with 0.2 mW microwave power, 100 kHz modulation frequency, 0.1 mT modulation amplitude and 10 ms conversion time and time constant.

DEER experiments were performed at 50 K on a Bruker ELEXSYS E580 spectrometer operating at 9 GHz equipped with an ER-4118-X-MS-3W resonator. The four-pulse DEER sequence was chosen with $\pi/2(\nu_{\text{obs}}) - \tau_1 - \pi(\nu_{\text{obs}}) - t - \pi(\nu_{\text{pump}}) - (\tau_1 + \tau_2 - t') - \pi(\nu_{\text{obs}}) - \tau_2 - \text{echo}$, where the observer pulse length was 16 ns for $\pi/2$ and 32 ns for π pulses. The pump pulse length was 12 ns, the long interpulse delay was $\tau_2 = 3 \mu$ s. All other parameters were used according to ref. 37. The DEER spectra were analysed using the programme DeerAnalysis2010 (ref. 38). The background was corrected by a homology three-dimensional fit. Simulations were checked for stability according to the DeerAnalysis2010 manual.

Functional analysis of FimD CTDs. The FimDACTD1+2, Δ CTD2 only, and D725R+N728R mutants were derived from plasmid pETS4 using the SLIM protocol⁴⁰ (Supplementary Table 3). The expression level of the FimD mutants in the outer membrane was similar to wild-type FimD. Ability of the mutants to assemble functional pili on the bacterial surface was determined by haemagglutination assay, carried out as described previously¹⁹.

- Chiu, J., Tillett, D., Dawes, I. W. & March, P. E. Site-directed, Ligase-Independent Mutagenesis (SLIM) for highly efficient mutagenesis of plasmids greater than 8 kb. *J. Microbiol. Methods* **73**, 195–198 (2008).
- Leslie, A. G. The integration of macromolecular diffraction data. *Acta Crystallogr. D* **62**, 48–57 (2006).
- Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
- Otwinowski, Z. & Minor, W. in *Methods in Enzymology* Vol. 276 (ed. Carter, C. W. Jr) Ch. 20, 307–326 (Elsevier, 1997).
- Vagin, A. A. & Isupov, M. N. Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Crystallogr. D* **57**, 1451–1456 (2001).

Principles of activation and permeation in an anion-selective Cys-loop receptor

Ryan E. Hibbs¹ & Eric Gouaux^{1,2}

Fast inhibitory neurotransmission is essential for nervous system function and is mediated by binding of inhibitory neurotransmitters to receptors of the Cys-loop family embedded in the membranes of neurons. Neurotransmitter binding triggers a conformational change in the receptor, opening an intrinsic chloride channel and thereby dampening neuronal excitability. Here we present the first three-dimensional structure, to our knowledge, of an inhibitory anion-selective Cys-loop receptor, the homopentameric *Caenorhabditis elegans* glutamate-gated chloride channel α (GluCl), at 3.3 Å resolution. The X-ray structure of the GluCl–Fab complex was determined with the allosteric agonist ivermectin and in additional structures with the endogenous neurotransmitter L-glutamate and the open-channel blocker picrotoxin. Ivermectin, used to treat river blindness, binds in the transmembrane domain of the receptor and stabilizes an open-pore conformation. Glutamate binds in the classical agonist site at subunit interfaces, and picrotoxin directly occludes the pore near its cytosolic base. GluCl provides a framework for understanding mechanisms of fast inhibitory neurotransmission and allosteric modulation of Cys-loop receptors.

Fast inhibitory neurotransmission modulates both the magnitude and duration of neuronal activity, occurs on a timescale of milliseconds, and involves the release of inhibitory neurotransmitters into the synapse and activation of the cognate ligand-gated ion channels. As demonstrated nearly 60 years ago¹, fast inhibitory neurotransmission leads to an increase in the permeability of the cell membrane to chloride, the most abundant biological anion. Because the membrane potential at which chloride is at equilibrium is near the neuronal resting potential, neurotransmitter-gated, chloride-selective ion channels generally oppose normal excitability and repolarize the cell².

The neurotransmitter receptors that directly mediate chloride permeability constitute one half of the Cys-loop receptor family³. Receptors in this family are composed of five either identical or homologous subunits, which generate diversity in functional profiles and pharmacological preferences. Cys-loop receptors fall into two broad categories. The cation-selective members are the nicotinic acetylcholine (nAChR) and serotonin 5-HT₃ receptors. Those selective for anions include the γ -aminobutyric acid (GABA_{A/C}), glycine receptors and invertebrate glutamate-gated chloride channels (GluCl)^{3–6}. So far, there is no structural information for an anion-selective Cys-loop receptor, and the mechanism by which chloride is selected remains unclear.

Ligand-gated chloride channels are critical not only for maintaining appropriate neuronal activity, but have long been important therapeutic targets: benzodiazepines, barbiturates, some intravenous and volatile anaesthetics, alcohol, strychnine, picrotoxin and ivermectin all derive their biological activity from acting on the inhibitory half of the Cys-loop receptor family^{3,7}. Of note is that many of the therapeutically useful compounds acting at Cys-loop receptors target an allosteric site. The sites in Cys-loop receptors at which these allosteric ligands bind and their structure-based mechanisms of action are largely unresolved.

Crystallization of GluCl–Fab complex

We identified the *Caenorhabditis elegans* GluCl α glutamate-gated chloride channel⁸ as a promising candidate using fluorescence-detection size-exclusion chromatography (FSEC)⁹. In comparison to human

Cys-loop receptors, GluCl α is most similar to the α 1 glycine receptor, with which it shares 34% amino acid sequence identity (see alignment in Supplementary Fig. 1). Optimization of the receptor construct for crystallization (GluCl_{cryst}) was guided by FSEC analysis and required deletion of 41 residues from the amino terminus, 6 residues from the carboxy terminus and replacement of the M3–M4 loop (Lys 345–Lys 402) with an Ala–Gly–Thr tripeptide. Well-ordered crystals diffracting to \sim 3.3 Å resolution required co-crystallization of GluCl_{cryst} as a complex with a Fab, ivermectin and lipids (Supplementary Fig. 2). Structures with agonist or channel blocker at 3.3 and 3.4 Å were obtained by soaking GluCl_{cryst}–Fab–ivermectin crystals with glutamate or picrotoxin, respectively. The electron density maps are of high quality, thus enabling the positioning of almost all receptor residues and refinement to satisfactory crystallographic residuals and stereochemistry (Supplementary Table 1 and Supplementary Fig. 3).

Architecture

The GluCl_{cryst}–Fab complex forms a pinwheel shape comprising a cylindrical homopentamer of GluCl_{cryst} subunits with Fab molecules bound at each subunit interface (Fig. 1a, b). Each GluCl_{cryst} subunit consists of a large N-terminal extracellular domain of mostly β structure, followed by four α -helical transmembrane spans (M1–M4; Fig. 1c). The overall architecture of the extracellular domain is similar to that found in the bacterial receptor orthologues from *Gloeobacter violaceus* (GLIC)^{10,11} and *Erwinia chrysanthemi* (ELIC)¹². There is an additional helix at the N terminus reminiscent of the acetylcholine-binding protein^{13–15} (AChBP) and *Torpedo marmorata* nAChR¹⁶ structures. Significantly, GluCl contains the Cys-loop disulphide strictly conserved in eukaryotes as well as a disulphide bond in loop C present in glycine receptors (Fig. 1c). The transmembrane helices adopt a fold like the bacterial receptors and nAChR, with the five M2 segments lining the pore and adopting an open channel conformation, akin to the conformation visualized in the GLIC structures.

To understand the molecular principles of ion channel activation, agonist binding and ion channel permeation and block, we determined

¹Vollum Institute, Oregon Health and Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239, USA. ²Howard Hughes Medical Institute, Oregon Health and Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239, USA.

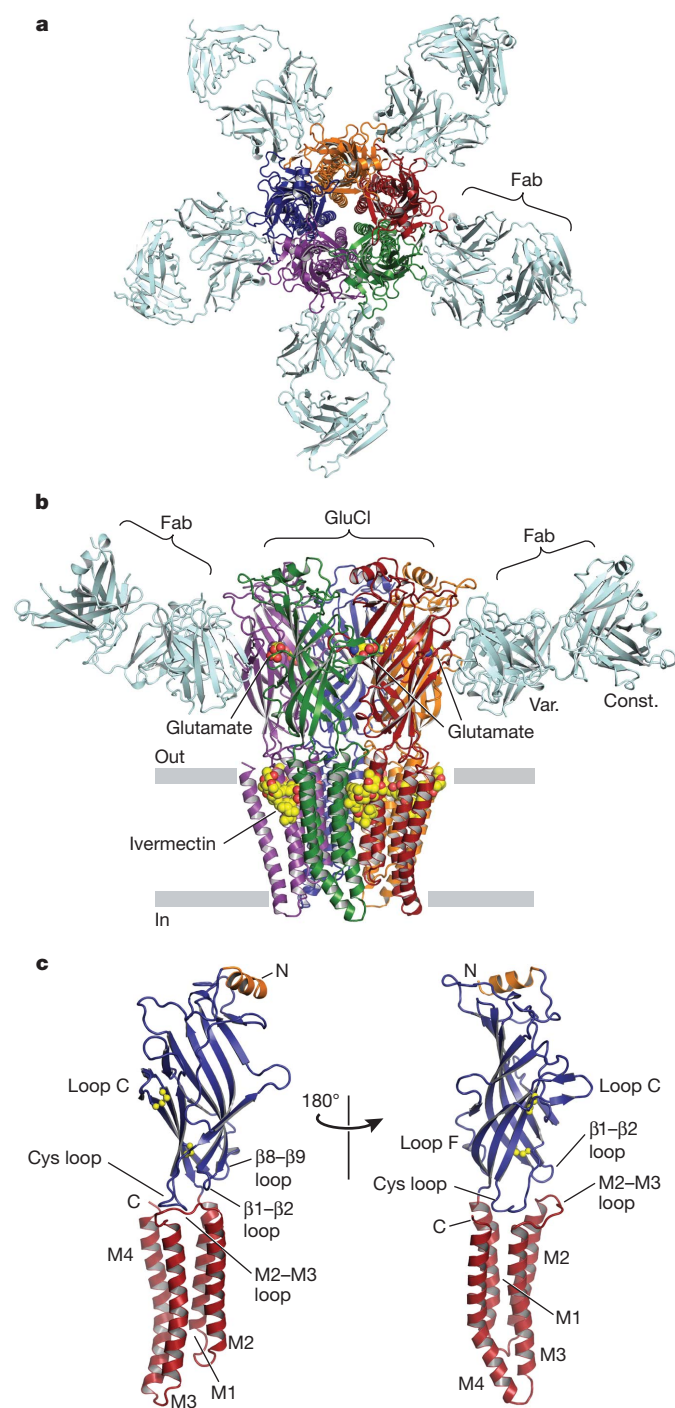


Figure 1 | Architecture of the GluCl_{cryst}-Fab complex. **a**, View of the GluCl_{cryst}-Fab complex looking down pore axis towards cytosol. Fab molecules (cyan) are bound at each GluCl_{cryst} subunit interface. **b**, View parallel to lipid membrane; only two Fab molecules are shown for clarity. The ligands ivermectin and glutamate are represented as spheres with carbon atoms in yellow, oxygen in red and nitrogen in blue. Const., constant region; Var., variable region. **c**, A single GluCl_{cryst} subunit from two angles, approximate orientation as in panel **b**. The Cys loop and loop C disulphide bonds are shown as spheres, N and C termini and transmembrane spans are indicated. Loops of particular relevance to agonist binding and allosteric gating linkage are also indicated.

separate crystal structures with the allosteric agonist ivermectin, and with ivermectin and glutamate, picrotoxin or iodide. Ivermectin is bound at each of the GluCl_{cryst} subunit interfaces in the transmembrane domain whereas glutamate electron density is present in all five of the classical neurotransmitter-binding sites in the extracellular

domain. Anomalous difference density for iodide is present in sites at the base of the transmembrane pore in a region important for ion selectivity, and a chloride ion was fit into non-protein electron density in the ion channel pore adjacent to the binding site for picrotoxin.

Allosteric activation and modulation

Ivermectin is a semi-synthetic macrocyclic lactone and broad-spectrum antiparasitic agent, widely used to treat river blindness in humans and parasitic infections in animals^{17,18}. It achieves its margin of therapeutic efficacy by activating invertebrate glutamate-gated chloride channels at nanomolar concentrations^{8,19}, yet it also manifests activating and potentiating activities on vertebrate Cys-loop receptors^{20–22} and on P2X ATP-gated ion channels²³ at higher concentrations. Ivermectin potentially activates GluCl α (Supplementary Fig. 4) while simultaneously rendering the receptor susceptible to further activation by glutamate²⁴. Hence, at GluCl α , we deem ivermectin a partial allosteric agonist.

Ivermectin binds at subunit interfaces on the periphery of the transmembrane domains, proximal to the extracellular side of the membrane bilayer (Fig. 2a, b and Supplementary Figs 5–7). Wedged between the M3 α -helix on the principal or (+) subunit and the M1 α -helix on the complementary or (–) subunit, ivermectin inserts deeply into the subunit interface and makes important contacts with the M2 (+) pore-lining α helix and the M2–M3 loop. Its site occupies approximately two turns of helix on the M1 and M3 helices and centres on a single turn of π helix between residues Leu 217 and Ile 222 on M1, as illustrated by a hydrogen bond between the main-chain carbonyl oxygen of Leu 218 and a tertiary hydroxyl on ivermectin (Fig. 2c). Through extensive hydrophobic interactions and one hydrogen bond with each of the M1, M2 and M3 α -helices, ivermectin buries 278 and 254 Å² of surface area on the (+) and (–) subunits in the interface, respectively.

Ser 260 forms a hydrogen bond with the secondary hydroxyl group on the deeply buried cyclohexene ring of ivermectin (Fig. 2a–c and Supplementary Fig. 8). A serine residue in this position is correlated with direct activation by ivermectin in other Cys-loop receptors. Glycine and GABA_A receptors have a serine in the equivalent position and are directly activated by ivermectin^{20,21}, yet there is no similar serine in $\alpha 7$ nAChRs, where ivermectin is a positive allosteric modulator but does not directly activate²², nor in GluCl β receptors, where ivermectin has no activity⁸. The equivalent position is critical for GABA_A receptor modulation by alcohol²⁵, anticonvulsants, anaesthetics and diuretics; glycine and 5-HT₃ receptor modulation by anaesthetics⁶; and $\alpha 7$ nAChR modulation by additional compounds²⁶. Hence, the ivermectin binding site in GluCl_{cryst} is shared, at least in part, by many important modulators of Cys-loop receptors. In GluCl we suspect that the interaction of ivermectin with the pore-lining M2 helix increases both its affinity for the receptor and its ability to stabilize the open state.

Ivermectin binding to GluCl probably results in two types of conformational changes: first, a local distortion of the receptor in the vicinity of the binding site; and second, a global conformational change of the receptor that corresponds to a transition from a closed, resting state to an open, activated state. Because we lack a structure of GluCl_{cryst} in the absence of ivermectin, GLIC provides a reference for gauging the local structural consequences of ivermectin binding to the transmembrane domain of the receptor. In comparing these two structures we find that the binding of ivermectin increases the separation between M1 and M3 of adjacent subunits, as defined by a 9.4 Å spacing between GluCl_{cryst} Leu 218 and Gly 281 C α atoms compared to a 6.4 Å spacing for the corresponding atoms in GLIC. This splaying apart of the transmembrane helices in GluCl_{cryst} occurs at the level of a strictly conserved proline residue in M1 that forms the C-terminal end of the short π helix (Supplementary Fig. 9).

We hypothesize that the global conformational change induced by ivermectin binding is rooted in the splaying apart of the M1 and M3 helices and the movement of the apical portion of M2 away from the

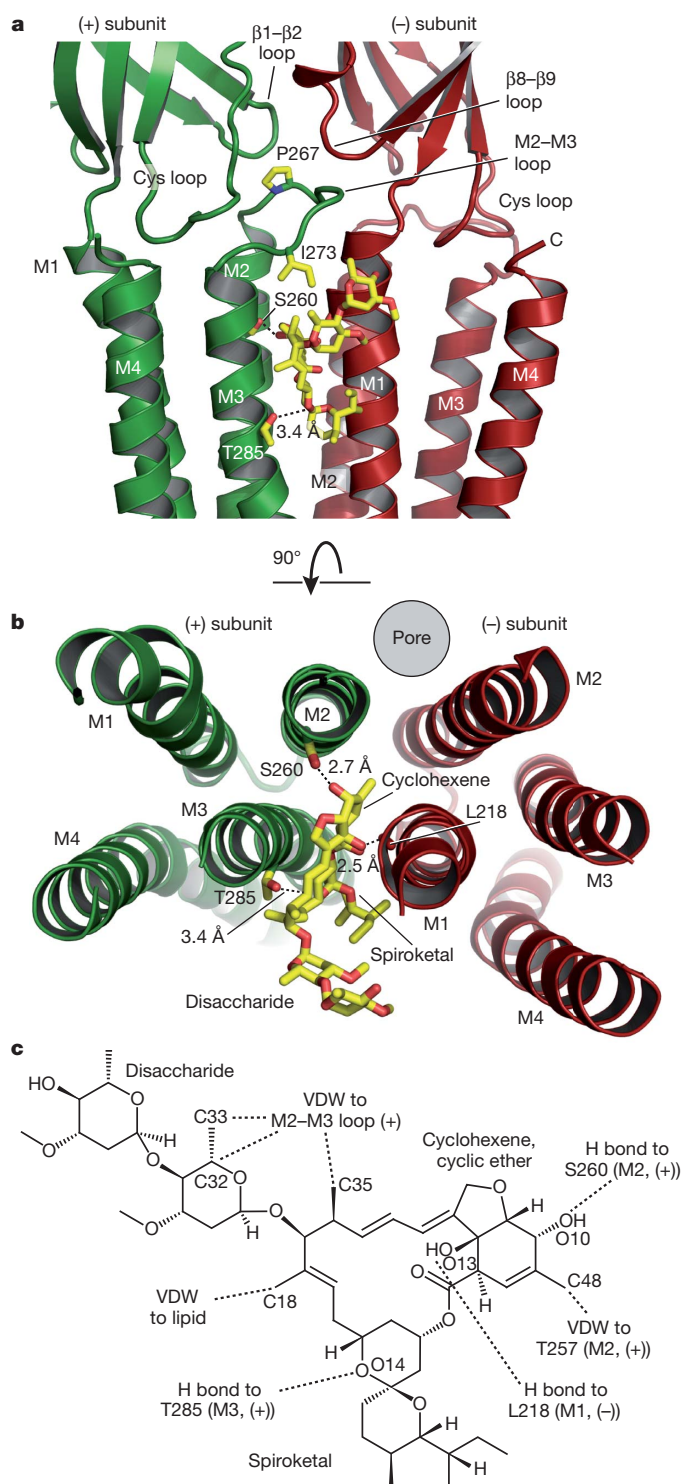


Figure 2 | Ivermectin-binding site and atomic interactions. **a**, **b**, Two orientations of a GluCl subunit interface focusing on ivermectin-binding site. Dashed lines indicate hydrogen bonds. In **a**, view is from receptor periphery looking parallel to the membrane, and in **b** looking down pore from extracellular side with the extracellular domain removed for clarity. **c**, Chemical structure of ivermectin with interactions indicated. VDW, van der Waals. Atomic numbering is from PDB file.

pore axis, towards the periphery of the receptor, opening an ion conductive pathway. This open-pore conformation of M2 is then stabilized through interactions between ivermectin and the apical end of M2. In addition, ivermectin may stabilize the open state of the ion channel through contacts between the disaccharide moiety and Ile 273 in the M2–M3 loop (Supplementary Fig. 10).

Neurotransmitter-binding site

The ion channel of GluCl_{cryst} is activated by glutamate only after activation by ivermectin (Supplementary Fig. 11), in a manner similar to full-length GluCl α (ref. 24). The homomeric GluCl β and heteromeric GluCl $\alpha\beta$ receptors, by contrast, are directly activated by glutamate⁸. In the context of GluCl_{cryst}, micromolar concentrations of glutamate augment ivermectin-induced currents by 30–70%, similar to that of full-length receptor. [³H]-L-glutamate binds directly to the GluCl_{cryst} receptor with a dissociation constant (K_d) of 680 nM (Supplementary Figs 11–13). In agreement with the electrophysiology experiments, [³H]-L-glutamate binding requires ivermectin. To understand the molecular basis of glutamate binding we determined the structure of GluCl_{cryst} in the presence of ivermectin and glutamate. Sausage-shaped electron density assigned to glutamate was $\sim 8\sigma$ in $F_o - F_c$ omit maps in all five of the classical agonist-binding sites. Omit electron density maps subjected to real space, five-fold averaging showed a protrusion in the electron density ‘sausage’ that we attributed to the α -amino group of glutamate (Supplementary Fig. 14).

Glutamate binds in the classical neurotransmitter site in the extracellular domain^{3,27}, lodged between subunits and nearly inaccessible to solvent (Fig. 3a, b). The architecture of the site is box-like, with loops from the (+) subunit forming ‘sides’ of the binding site and the β strands on the (–) subunit defining the ‘base.’ Loop C, postulated to have a critical role in allosteric activation^{15,28–30}, adopts a closed conformation consistent with AChBP structures bound by agonists. Functional groups on glutamate bridge the (+) and (–) subunits with the α -substituents snugly sandwiched between Tyr 151 and Tyr 200 on the (+) subunit, and positively charged residues, including Arg 37 from a region important for conotoxin–nAChR interaction³¹, and Arg 56 on the (–) subunit, making contacts with the α - and γ -carboxylate groups. These arginine residues, in combination with neighbouring cationic amino acids, provide the binding pocket with a strongly positive electrostatic potential (Supplementary Fig. 15). The α -amino nitrogen of glutamate is stabilized through a 3.8 Å cation– π interaction with Tyr 200 on loop C, a hydrogen bond with the backbone carbonyl oxygen of Ser 150 and a close interaction with the backbone carbonyl oxygen of Tyr 151. A comparison of the determinants of glutamate binding with the corresponding residues in the AChBPs and other receptors is made in Supplementary Fig. 16.

To test the sensitivity of the glutamate binding site to perturbations in ligand structure, we screened glutamate analogues for competition with [³H]-L-glutamate bound to the ivermectin-complexed receptor (Fig. 3c and Supplementary Fig. 17). L-Glutamate bound much tighter than L-homocysteine sulphonic acid, which differs only in replacement of C δ with sulphur. Extending the side-chain length with an extra carbon (L-amino adipic acid) or shortening it (L-aspartate) resulted in a further drop in affinity, and changing the stereochemistry (D-glutamate) or removing the side-chain negative charge but not its ability to hydrogen bond (L-glutamine) decreased binding further. Thus, the GluCl neurotransmitter-binding pocket is selective for small dicarboxylate L-amino acids, consistent with the constellation of atomic interactions between agonist and receptor (Fig. 3a, b).

Upon the binding of glutamate the side chain of Arg 56 in loop D ($\beta 2$) shifts by ~ 0.5 Å to accommodate the agonist and Tyr 200 in loop C repositions by ~ 0.5 Å closer to the ligand, small yet significant conformational changes consistent with movements of loops C and D in the agonist-induced activation of the receptor. These residues, together with Arg 37 ($\beta 1$), are located on elements of protein structure directly connected to the ion channel pore. We suggest that ivermectin, a partial allosteric agonist, stabilizes an ‘activated’ conformation of the agonist site and that binding of glutamate to this ‘activated’ site further stabilizes the open state of the receptor, increasing chloride conductance. Ivermectin may transduce a conformational change to the neurotransmitter site through its interactions with the M2–M3 loop, located at the structural nexus of three extracellular domain loops central to allosteric communication between the neurotransmitter site in the extracellular domain and the transmembrane pore: the Cys,

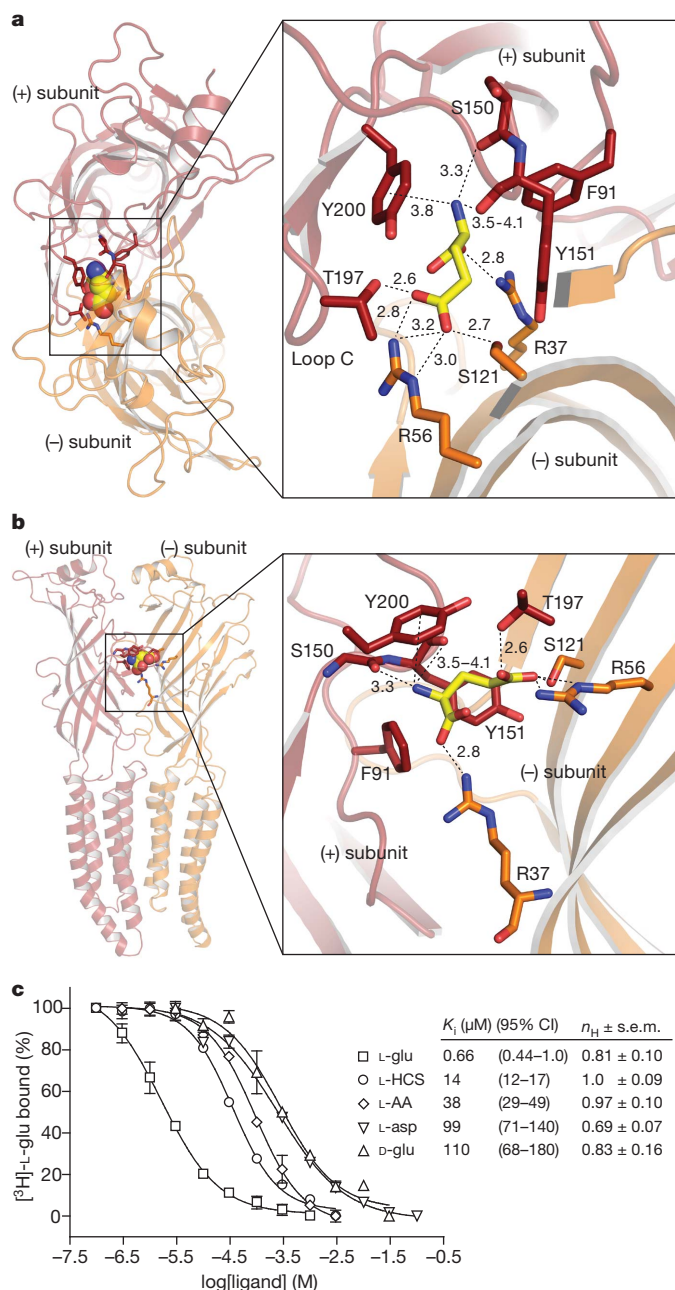


Figure 3 | Glutamate-binding site and specificity. **a**, View from extracellular side towards membrane at glutamate in binding site in subunit interface. **b**, View of binding site looking parallel to membrane with loop C removed for clarity. Dashed lines with distances in Å indicate hydrogen bonding and, in the case of Tyr 200, cation- π interactions. Unless a range is given, distances are an average from the five binding sites. **c**, Radioligand competition experiments with L-glutamate and congeners against 1 mM [^3H]-L-glutamate. Calculated inhibition constant (K_i) values assume a K_d for [^3H]-L-glutamate of 680 nM and are shown in inset table. $n = 2$. CI, confidence interval. L-HCS and L-AA are L-homocysteine sulphinic acid and L-amino adipic acid, respectively. Error bars are s.e.m. and n_H is the Hill coefficient.

$\beta 1$ – $\beta 2$ and $\beta 8$ – $\beta 9$ loops⁶ (Fig. 2a and Supplementary Fig. 10). Hydrophobic residues in the M2–M3 loop mediating these interactions are well conserved in most Cys-loop receptors, consistent with the M2–M3 loop having a central role in the activation mechanism of receptors throughout the family^{32–36}.

Pore conformation

To test the hypothesis that the GluCl_{cryst}–ivermectin structure represents an open, conducting conformation (Fig. 4), we carried out

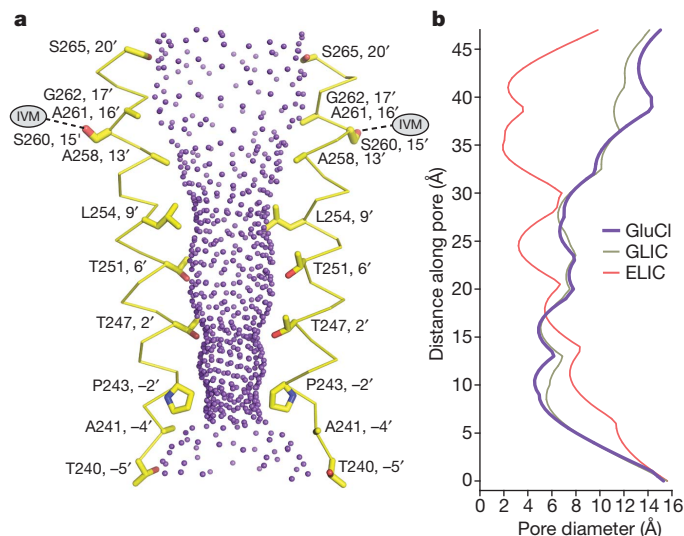


Figure 4 | Ion channel. **a**, Purple spheres represent internal surface of transmembrane ion channel, with side chains shown for pore-lining residues from two of the five M2 α helices that line the pore; Ser 260 does not line the pore but hydrogen bonds with ivermectin (IVM). **b**, Pore diameter is plotted as a function of longitudinal distance along the pore for GluCl_{cryst}, open (GLIC; PDB code 3EAM) and closed (ELIC; PDB code 2VL0) bacterial receptors.

functional and structural studies using picrotoxin, an open channel blocker^{37,38} (Fig. 5 and Supplementary Fig. 18). Electron density in picrotoxin-soaked crystals was apparent at a position near the cytosolic side of the transmembrane pore, on the five-fold axis of molecular symmetry (4.3σ ; Supplementary Fig. 19). Thus, the observed electron density is an average of five orientations. Nevertheless, the egg-shaped picrotoxin-associated electron density indicates that the basket-like, fused tricyclic rings are directed extracellularly and near the 2' Thr, whereas the isoprenyl tail points towards the cytoplasm and is proximal to the –2' Pro residues. In this position, the majority of the oxygen atoms of picrotoxin are cradled by the polar belt of 2' Thr hydroxyls whereas the hydrophobic isoprenyl moiety is surrounded by the methylene groups of the non-polar –2' Pro side chains. Most importantly, the binding of picrotoxin to the pore of the GluCl_{cryst}–ivermectin complex reinforces our hypothesis that the pore is in an open conformation.

The smallest diameter of the GluCl_{cryst} ion channel pore is ~ 4.6 Å, defined by a hydrophobic ‘girdle’ of –2' Pro side chains proximal to the cytoplasmic side of the membrane. Because chloride has a Pauling

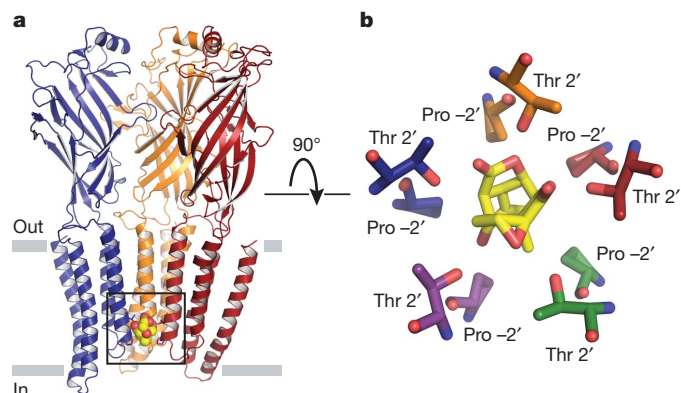


Figure 5 | Picrotoxin-binding site. **a**, The front two subunits have been removed to show the picrotoxin location (boxed) at the cytosolic base of the pore. Residues involved in picrotoxin binding are shown as sticks and van der Waals surfaces are shown for picrotoxin. **b**, View looking into pore from the extracellular domain at the picrotoxin position relative to the 2' Thr and –2' Pro side chains. Picrotoxin is shown as sticks with carbon atoms in yellow and oxygen atoms in red.

radius of 1.8 Å (ref. 2), passage of chloride, iodide (Pauling radius of 2.2 Å (ref. 2)) and other permeant ions through the $-2'$ Pro constriction must involve substantial dehydration, in agreement with previous studies demonstrating a correlation between energies of hydration and relative permeabilities (higher for iodide than chloride; Supplementary Fig. 20)³⁹. The pore constriction in GluCl_{cryst} is somewhat smaller than that estimated for GABA_A, glycine and GluCl β receptors (5.2–6.2 Å)^{3,40}, based on low but measureable relative permeability to ions like acetate, gluconate and phosphate. This difference may be due to the alanine residues at the $-2'$ position in the β subunits of all three of those receptors.

GluCl is related to the *Torpedo* nAChR (PDB code 2BG9)¹⁶ in amino acid sequence and three-dimensional structure and thus we compared the structures and the aligned sequences. In so doing, we found inconsistencies between amino acid sequence-based alignments and three-dimensional structure-based alignments of the M2 and M3 α helices. A similar finding was described in comparisons of GLIC to the nAChR⁴. Our analysis indicates that in the α subunit M2 pore-lining helix and the M3 α -helix, the nAChR amino acid assignment is off in register by 4 residues or ~ 1 turn of an α -helix beginning with the M1–M2 loop (Supplementary Fig. 21).

Ion selectivity

Analysis of GluCl_{cryst} surface electrostatics reveals an electropositive vestibule, a slightly electronegative extracellular half of the transmembrane pore, and an electropositive intracellular half (Fig. 6a). None of the pore-lining residues in GluCl_{cryst} bear a formal charge and thus the positive electrostatic potential at the base of the pore arises from the oriented peptide dipoles in the M2 α helices⁴¹, reminiscent of the role that helical dipoles have in ClC chloride channels⁴². Cation channels

reverse the selectivity imposed by orientation of the M2 dipoles through placing a negatively charged side chain near the pore constriction point⁴³. Although other regions contribute to the modulation of conductance and selectivity in some Cys-loop receptors, the minimal determinants of selectivity are the $-1'$ Ala and $-2'$ Pro positions for anions and the $-1'$ Glu for cations^{3,44}, with no requirement for positively charged amino acids in the pore of anion-selective channels⁴⁰ (Supplementary Fig. 22).

To identify sites important in chloride binding and selectivity, we soaked crystals of GluCl_{cryst} in iodide, a heavy atom analogue of chloride, and observed four anomalous difference peaks that we ascribe to iodide, located at the cytosolic base of the transmembrane pore and centred around the five-fold symmetry axis (Fig. 6b–d). The weak density at the fifth site is simply the consequence of an interfering lattice contact with an adjacent Fab. Each iodide sits in a concave pocket of positive electrostatic potential formed by $-2'$ Pro residues from the M2 helices of adjacent subunits, main-chain backbone atoms of $-1'$ Ala and $-3'$ Ile and the methyl group of $-1'$ Ala. All three of these residues are important in selectivity for some receptors, with the $-1'$ position being an essential component of selectivity across the family³. Previous studies indicate that the main chain amide nitrogen at the $-3'$ position is important in GluCl β receptors for anion dehydration⁴⁰. In GluCl_{cryst} this atom is ~ 5 Å from the centre of the iodide anomalous density and could form water-mediated hydrogen bonds to anions at the mouth of the ion channel pore.

Electron density maps derived from all GluCl_{cryst} X-ray diffraction data sets exhibit a spherical peak in the pore between the $2'$ Thr and $6'$ Thr residues (6.8σ in $F_o - F_c$ omit maps) with no other peaks in the pore above 2.5σ . Anomalous difference electron density maps were inconclusive in identification of this peak. Therefore, we placed

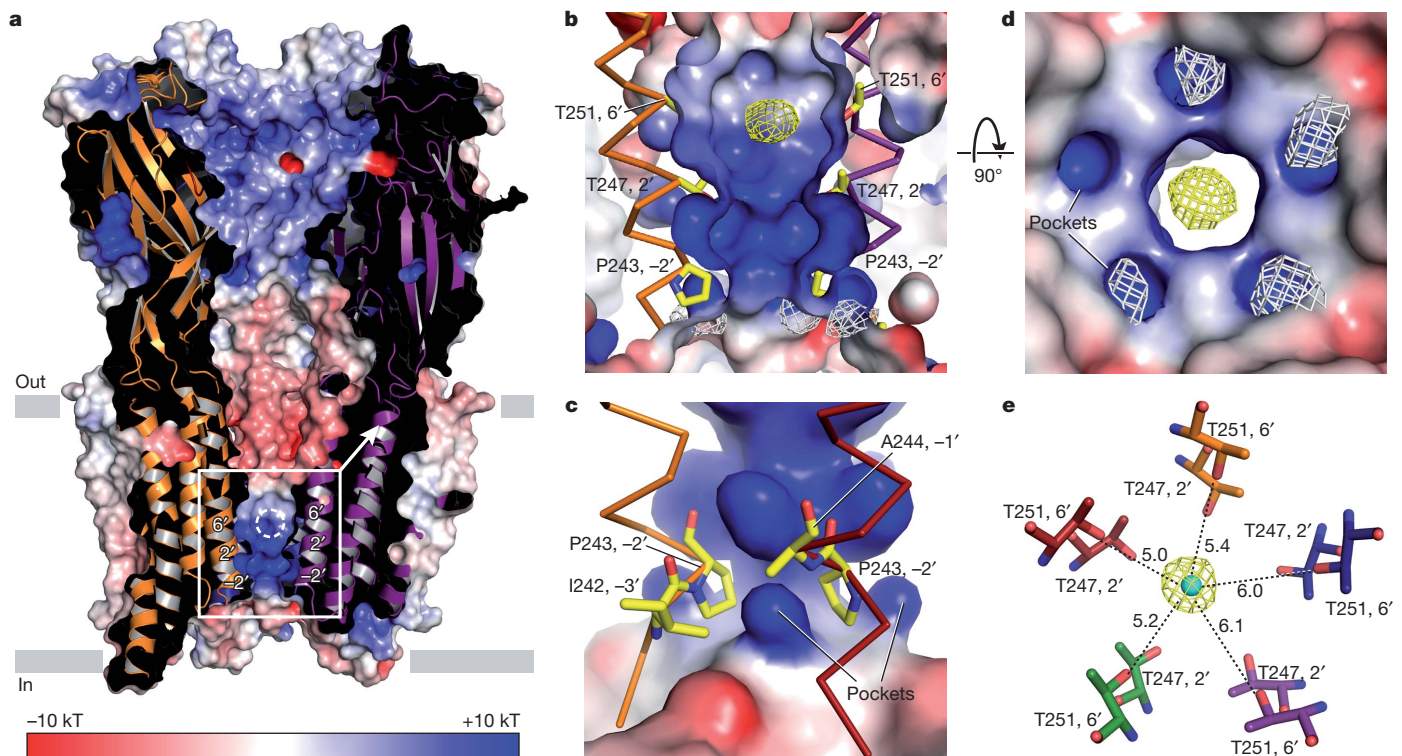


Figure 6 | Ion selectivity. **a**, The front of the receptor is cut away to reveal the interior surface of the pore, coloured by electrostatic potential. Dashed circle in the pore indicates putative chloride-binding site. Boxed area is expanded in **b**. **b**, Expanded view from **a** showing selected M2 side chains from opposing subunits. Anomalous difference peaks at pore base are attributed to iodide-binding sites (light grey mesh, contoured at 3.5σ). $F_o - F_c$ omit density for the putative chloride site is represented by yellow mesh contoured at 3σ . **c**, Shown are the electropositive pockets where iodide ions bind, viewed from inside the

protein surface; four residues from adjacent M2 helices that coordinate the iodide sites are shown as sticks. **d**, Electropositive pockets viewed from the intracellular side. **e**, Putative chloride site viewed from the extracellular side of the pore with $6'$ Thr residues in the foreground; $F_o - F_c$ omit density (yellow) is contoured at 4σ . Carbon atoms are coloured by chain and chloride is represented by a 1 Å cyan sphere. Closest distances in Å from protein atoms to centre of sphere are indicated by dashed lines from the $6'$ Thr side-chain hydroxyl.

several different ions or molecules in the difference density and determined, through crystallographic refinement and inspection of difference electron density maps, that a single chloride anion best accounted for this electron density feature (Supplementary Table 1). Distances between the modelled chloride and 6' Thr side-chain hydroxyl oxygen atoms are consistent with water-mediated hydrogen bonding of chloride in the pore⁴⁵, indicating that this location could be a transiently occupied ion-binding site flanking the constriction point. Further experimentation is required to validate the chemical identity of the bound species.

The iodide-binding sites nestled in electropositive pockets at the base of the pore suggest general principles of ion selectivity in Cys-loop receptors. In GluCl and other chloride-selective receptors there is either an alanine or glycine residue at the $-1'$ position of the M2 helix, thus preserving the concave pocket. By contrast, in eukaryotic cation-selective channels, the $-1'$ residue is a conserved glutamate. We suggest that the carboxylate side chain of glutamate not only fills the 'anion pocket' but that it also imposes a local negative electrostatic potential important for cation selectivity (Supplementary Fig. 22). Previous cysteine accessibility studies in cation-selective channels have indicated that the $-1'$ Glu position lines the transmembrane pore⁴⁶. However, on the basis of the GluCl_{cryst} structure and amino acid sequence alignments, we propose that the preceding residue, a conserved glycine ($-2'$ residue), lines the pore of cation channels, consistent with the significantly larger pore diameter of cation channels (7.4–8.4 Å). In support of the $-2'$ residue defining the pore constriction, deletion of the $-2'$ Pro in glycine receptors, which would shift the following glycine residue into the $-2'$ position, increases pore diameter to 6.9 Å (ref. 47). Furthermore, the $-2'$ Gly position in cation-selective 5-HT_{3A} receptors is accessible to modification when the pore is open⁴⁸. We propose that the $-2'$ position lines the pore in both anion and cation channels and that the 'anion pockets' in GluCl_{cryst} are important determinants of ion selectivity, increasing the local concentration of anions at the cytoplasmic mouth of the pore.

Conclusion

Here we present the first X-ray structure, to our knowledge, of a eukaryotic Cys-loop receptor, a glutamate-gated chloride channel from *C. elegans*. GluCl_{cryst} was co-crystallized with ivermectin, a partial allosteric agonist that sequesters within the membrane bilayer and binds to exposed sites on the transmembrane domains of the receptor. Lipophilic modulators of other Cys-loop receptors may exploit a similar mechanism of interaction, including the neurosteroids at the GABA_A receptor⁶ and cholesterol at the muscle nAChR⁴⁹. The GluCl_{cryst}–ivermectin structure maps a previously uncharacterized binding site at a protein–lipid interface and defines a protein/chemical scaffold for design of receptors and ligands with new pharmacological properties and receptor specificities. Binding of ivermectin induces local changes in the membrane domain and global conformational changes in the entire receptor, pre-organizing the agonist binding site ~30 Å away and opening the ion channel pore. Analysis of amino acids lining and proximal to the pore indicates that anion selectivity is accomplished largely through a pore constriction imposed by proline residues and a positive electrostatic potential, conferred by the N-terminal end of the M2 helix dipoles. These new findings advance our understanding of the molecular mechanism of fast neuronal inhibition, the importance of which was first appreciated more than one hundred years ago⁵⁰.

METHODS SUMMARY

GluCl_{cryst} was expressed from baculovirus-infected Sf9 cells and purified by metal ion affinity chromatography. The Fab complex was isolated by size-exclusion chromatography. The GluCl_{cryst}–Fab complex was concentrated to 1–2 mg ml⁻¹ and supplemented with synthetic lipids and ivermectin. Crystallization was performed by hanging-drop vapour diffusion at 4 °C with a precipitating solution containing 21–23% PEG 400, 50 mM sodium citrate pH 4.5 and 70 mM sodium

chloride. Cryoprotection was achieved by soaking crystals in precipitant solution supplemented with 30% PEG 400. Additional complexes were obtained by soaking crystals in cryoprotectant containing L-glutamate, picrotoxin or sodium iodide. Diffraction data were indexed, integrated and scaled and the structure solved by molecular replacement using a GLIC-derived homology model of GluCl_{cryst} and a Fab homology model as search probes. The molecular replacement phases were used to initiate autobuilding and the resulting model was iteratively improved by cycles of manual adjustment and crystallographic refinement. Function of GluCl was examined by two-electrode voltage clamp experiments and by [³H]-L-glutamate saturation and competition binding assays.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 March; accepted 21 April 2011.

Published online 15 May 2011.

- Coombs, J. S., Eccles, J. C. & Fatt, P. The specific ionic conductances and the ionic movements across the motoneuronal membrane that produce the inhibitory post-synaptic potential. *J. Physiol. (Lond.)* **130**, 326–374 (1955).
- Hille, B. *Ion Channels of Excitable Membranes* (Sinauer Associates, 2001).
- Thompson, A. J., Lester, H. A. & Lummis, S. C. The structural basis of function in Cys-loop receptors. *Q. Rev. Biophys.* **43**, 449–499 (2010).
- Corringer, P. J. et al. Atomic structure and dynamics of pentameric ligand-gated ion channels: new insight from bacterial homologues. *J. Physiol. (Lond.)* **588**, 565–572 (2010).
- Hilf, R. J. & Dutzler, R. A prokaryotic perspective on pentameric ligand-gated ion channel structure. *Curr. Opin. Struct. Biol.* **19**, 418–424 (2009).
- Miller, P. S. & Smart, T. G. Binding, activation and modulation of Cys-loop receptors. *Trends Pharmacol. Sci.* **31**, 161–174 (2010).
- Garcia, P. S., Kolesky, S. E. & Jenkins, A. General anesthetic actions on GABA_A receptors. *Curr. Neuropharmacol.* **8**, 2–9 (2010).
- Cully, D. F. et al. Cloning of an avermectin-sensitive glutamate-gated chloride channel from *Caenorhabditis elegans*. *Nature* **371**, 707–711 (1994).
- Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).
- Hilf, R. J. & Dutzler, R. Structure of a potentially open state of a proton-activated pentameric ligand-gated ion channel. *Nature* **457**, 115–118 (2009).
- Bocquet, N. et al. X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation. *Nature* **457**, 111–114 (2009).
- Hilf, R. J. & Dutzler, R. X-ray structure of a prokaryotic pentameric ligand-gated ion channel. *Nature* **452**, 375–379 (2008).
- Brejck, K. et al. Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411**, 269–276 (2001).
- Celie, P. H. et al. Crystal structure of acetylcholine-binding protein from *Bulinus truncatus* reveals the conserved structural scaffold and sites of variation in nicotinic acetylcholine receptors. *J. Biol. Chem.* **280**, 26457–26466 (2005).
- Hansen, S. B. et al. Structures of *Aplysia* AChBP complexes with nicotinic agonists and antagonists reveal distinctive binding interfaces and conformations. *EMBO J.* **24**, 3635–3646 (2005).
- Unwin, N. Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. *J. Mol. Biol.* **346**, 967–989 (2005).
- Campbell, W. C., Fisher, M. H., Stapley, E. O., Albers-Schonberg, G. & Jacob, T. A. Ivermectin: a potent new antiparasitic agent. *Science* **221**, 823–828 (1983).
- Aziz, M. A., Diallo, S., Diop, I. M., Lariviere, M. & Porta, M. Efficacy and tolerance of ivermectin in human onchocerciasis. *Lancet* **320**, 171–173 (1982).
- Arena, J. P., Liu, K. K., Paress, P. S. & Cully, D. F. Avermectin-sensitive chloride currents induced by *Caenorhabditis elegans* RNA in *Xenopus* oocytes. *Mol. Pharmacol.* **40**, 368–374 (1991).
- Adelsberger, H., Lepier, A. & Dudel, J. Activation of rat recombinant $\alpha_1\beta_2\gamma_2\delta$ GABA_A receptor by the insecticide ivermectin. *Eur. J. Pharmacol.* **394**, 163–170 (2000).
- Shan, Q., Haddrill, J. L. & Lynch, J. W. Ivermectin, an unconventional agonist of the glycine receptor chloride channel. *J. Biol. Chem.* **276**, 12556–12564 (2001).
- Krause, R. M. et al. Ivermectin: a positive allosteric effector of the α_7 neuronal nicotinic acetylcholine receptor. *Mol. Pharmacol.* **53**, 283–294 (1998).
- Silberberg, S. D., Li, M. & Swartz, K. J. Ivermectin interaction with transmembrane helices reveals widespread rearrangements during opening of P2X receptor channels. *Neuron* **54**, 263–274 (2007).
- Etter, A., Cully, D. F., Schaeffer, J. M., Liu, K. K. & Arena, J. P. An amino acid substitution in the pore region of a glutamate-gated chloride channel enables the coupling of ligand binding to channel gating. *J. Biol. Chem.* **271**, 16035–16039 (1996).
- Ueno, S., Wick, M. J., Ye, Q., Harrison, N. L. & Harris, R. A. Subunit mutations affect ethanol actions on GABA_A receptors expressed in *Xenopus* oocytes. *Br. J. Pharmacol.* **127**, 377–382 (1999).
- Young, G. T., Zwart, R., Walker, A. S., Sher, E. & Millar, N. S. Potentiation of α_7 nicotinic acetylcholine receptors via an allosteric transmembrane site. *Proc. Natl Acad. Sci. USA* **105**, 14686–14691 (2008).
- Kao, P. N. et al. Identification of the α subunit half-cystine specifically labeled by an affinity reagent for the acetylcholine receptor binding site. *J. Biol. Chem.* **259**, 11662–11665 (1984).
- Darmle, V. N. & Karlin, A. Effects of agonists and antagonists on the reactivity of the binding site disulfide in acetylcholine receptor from *Torpedo californica*. *Biochemistry* **19**, 3924–3932 (1980).

29. Celie, P. H. *et al.* Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron* **41**, 907–914 (2004).
30. Mukhtasimova, N., Free, C. & Sine, S. M. Initial coupling of binding to gating mediated by conserved residues in the muscle nicotinic receptor. *J. Gen. Physiol.* **126**, 23–39 (2005).
31. Quiram, P. A., McIntosh, J. M. & Sine, S. M. Pairwise interactions between neuronal $\alpha 7$ acetylcholine receptors and α -conotoxin PnIB. *J. Biol. Chem.* **275**, 4889–4896 (2000).
32. Lee, W. Y. & Sine, S. M. Principal pathway coupling agonist binding to channel gating in nicotinic receptors. *Nature* **438**, 243–247 (2005).
33. Campos-Caro, A. *et al.* A single residue in the M2–M3 loop is a major determinant of coupling between binding and gating in neuronal nicotinic receptors. *Proc. Natl Acad. Sci. USA* **93**, 6118–6123 (1996).
34. Kusama, T., Wang, J. B., Spiwak, C. E. & Uhl, G. R. Mutagenesis of the GABA rho 1 receptor alters agonist affinity and channel gating. *Neuroreport* **5**, 1209–1212 (1994).
35. Lynch, J. W., Rajendra, S., Barry, P. H. & Schofield, P. R. Mutations affecting the glycine receptor agonist transduction mechanism convert the competitive antagonist, picrotoxin, into an allosteric potentiator. *J. Biol. Chem.* **270**, 13799–13806 (1995).
36. Rajendra, S. *et al.* Mutation of an arginine residue in the human glycine receptor transforms β -alanine and taurine from agonists into competitive antagonists. *Neuron* **14**, 169–175 (1995).
37. Takeuchi, A. & Takeuchi, N. A study of the action of picrotoxin on the inhibitory neuromuscular junction of the crayfish. *J. Physiol. (Lond.)* **205**, 377–391 (1969).
38. Etter, A. *et al.* Picrotoxin blockade of invertebrate glutamate-gated chloride channels: subunit dependence and evidence for binding within the pore. *J. Neurochem.* **72**, 318–326 (1999).
39. Fatima-Shad, K. & Barry, P. H. Anion permeation in GABA- and glycine-gated channels of mammalian cultured hippocampal neurons. *Proc. R. Soc. Lond. B* **253**, 69–75 (1993).
40. Sunesen, M. *et al.* Mechanism of Cl⁻ selection by a glutamate-gated chloride (GluCl) receptor revealed through mutations in the selectivity filter. *J. Biol. Chem.* **281**, 14875–14881 (2006).
41. Wada, A. The α -helix as an electric macro-dipole. *Adv. Biophys.* 1–63 (1976).
42. Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* **415**, 287–294 (2002).
43. Wilson, G. G., Pascual, J. M., Brooijmans, N., Murray, D. & Karlin, A. The intrinsic electrostatic potential and the intermediate ring of charge in the acetylcholine receptor channel. *J. Gen. Physiol.* **115**, 93–106 (2000).
44. Keramidas, A., Moorhouse, A. J., Schofield, P. R. & Barry, P. H. Ligand-gated ion channels: mechanisms underlying ion selectivity. *Prog. Biophys. Mol. Biol.* **86**, 161–204 (2004).
45. Mancinelli, R., Botti, A., Bruni, F., Ricci, M. A. & Soper, A. K. Hydration of sodium, potassium, and chloride ions in solution and the concept of structure maker/breaker. *J. Phys. Chem. B* **111**, 13570–13577 (2007).
46. Akabas, M. H., Kaufmann, C., Archdeacon, P. & Karlin, A. Identification of acetylcholine receptor channel-lining residues in the entire M2 segment of the α subunit. *Neuron* **13**, 919–927 (1994).
47. Lee, D. J., Keramidas, A., Moorhouse, A. J., Schofield, P. R. & Barry, P. H. The contribution of proline 250 (P-2') to pore diameter and ion selectivity in the human glycine receptor channel. *Neurosci. Lett.* **351**, 196–200 (2003).
48. Reeves, D. C., Goren, E. N., Akabas, M. H. & Lummis, S. C. Structural and electrostatic properties of the 5-HT₃ receptor pore revealed by substituted cysteine accessibility mutagenesis. *J. Biol. Chem.* **276**, 42035–42042 (2001).
49. Barrantes, F. J. Structural basis for lipid modulation of nicotinic acetylcholine receptor function. *Brain Res. Rev.* **47**, 71–95 (2004).
50. Sherrington, C. S. *Integrative Action of the Nervous System* (Yale Univ. Press, 1906).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to H. Lester for providing the initial GluCl construct, to D. Cawley for monoclonal antibody production, to J. Michel for Fab fragment cloning and sequencing, to C. Alexander and D. C. Dawson for providing *Xenopus* oocytes, to M. Mayer for advice and equipment related to oocyte experiments, and to L. Vaskalis for help with illustrations. We thank the staff at the Advanced Photon Source beamline 24-ID-C for assistance with X-ray data collection. We are particularly appreciative of discussions with E.G. laboratory members and E. McCleskey. This work was supported by an individual NIH National Research Service Award (F32NS061404) to R.E.H. E.G. is an investigator with the Howard Hughes Medical Institute.

Author Contributions R.E.H. and E.G. contributed to all aspects of the project.

Author Information Atomic coordinates and structure factors have been deposited with the Protein Data Bank under codes 3RHW, 3RIF, 3RI5 and 3RIA for the GluCl–Fab–ivermectin complex alone and with glutamate, with picrotoxin and with iodide, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to E.G. (gouauxe@ohsu.edu).

METHODS

Construct design. The gene encoding the full-length *C. elegans* GluCl α protein (Genbank accession code AAA50785.1)⁸, including the native signal peptide and a C-terminal 8 \times -histidine tag, was codon optimized and subcloned into the pFastBac1 vector for baculovirus-driven expression in Sf9 insect cells. A construct for FSEC-based⁹ small-scale screening of detergent stability, mutagenesis and purification additionally contained the enhanced GFP (EGFP)-coding sequence inserted into the M3–M4 loop region as previously described^{51–53}. To improve crystallization behaviour, 41 amino acid residues from the N terminus and 6 from the C terminus were removed, and residues K345–K402 (in the mature, full-length sequence), corresponding to the M3–M4 loop, were substituted with the residues AGT.

GluCl expression and purification. Bacmid and baculovirus were generated from pFastBac1 constructs and Sf9 cells were infected at 27 °C using standard methods. After 18 h of infection, cells were maintained shaking at 20 °C, and then harvested for purification after 72–96 h. Cells were collected by centrifugation at 6,200g and disrupted using an EmulsiFlex-C5 (Avestin) in buffer containing 20 mM Tris pH 7.4, 150 mM NaCl (TBS buffer), and 1 mM PMSF. The homogenate was clarified by centrifugation at 9,700g, and crude membranes were collected from the light membrane fraction by centrifugation at 125,000g. The membranes were mechanically homogenized and solubilized in 0.25 g C₁₂M (*n*-dodecyl- β -D-maltopyranoside; Anatrace) per gram of membranes in TBS. Solubilized membranes were centrifuged at 125,000g. Supernatant containing GluCl_{cryst} was bound to TALON Co²⁺-affinity resin (Clontech), washed with TBS solution containing 1 mM C₁₂M and 25 mM imidazole, and eluted with 250 mM imidazole. All purification steps were performed at 4 °C.

Monoclonal antibody generation and Fab purification. The mouse monoclonal antibody against GluCl (IgG1, λ) was obtained using standard methods⁵⁴. Specificity of the antibody for properly folded pentameric GluCl_{cryst} was assayed by FSEC and western blot. Cloning and sequencing of Fab antibody regions were performed from mouse hybridoma cells. Antibody was purified from hybridoma supernatants by cation exchange and protein A affinity chromatography. Fab fragments were generated by papain digest of whole antibody, and purified by protein A chromatography to remove Fc molecules and undigested material, followed by anion exchange.

Purification of GluCl_{cryst}–Fab complex. Eluent from Co²⁺-affinity purification and Fab from ion exchange were mixed to an excess of Fab to GluCl_{cryst} subunits, concentrated, and applied to a gel filtration column (Superose 6 10/300 GL, GE Healthcare Life Sciences) equilibrated in TBS + 1 mM C₁₂M. GluCl_{cryst}–Fab complex was concentrated to 1–2 mg ml^{−1}. For samples used in crystallization, 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) or 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC) lipids (Avanti Polar Lipids) were added to 0.02% from a 2% stock suspension in 20% DMSO, 80% gel filtration buffer, and ivermectin (Sigma) was added to 0.1 mM from a 10 mM stock in DMSO.

Crystallization and cryoprotection. Initial crystallization attempts of GluCl constructs in the absence of Fab resulted in poorly diffracting (8 Å) crystals that grew in a very limited range of crystallization conditions. Crystallization of the Fab complex occurred in diverse conditions; best diffracting crystals were obtained in hanging-drop format and diffracted to 4–5 Å. Crystals diffracting beyond 4 Å were obtained only in the presence of Fab, either POPC or DPPC, and ivermectin. These tetragonal crystals grew by vapour diffusion at 4 °C in 21–23% PEG 400, 50 mM sodium citrate pH 4.5, and 70 mM sodium chloride, and diffracted maximally to Bragg spacings of 3.26 Å (Supplementary Table 1). Crystals were protected before flash freezing in liquid nitrogen by 1–2 min soaks in crystallization solution supplemented to contain 30% PEG 400. To obtain structures of GluCl_{cryst} in complex with additional ligands, crystals of the same form were soaked briefly in cryoprotectant containing either 5 mM picrotoxin (picrotoxinin, the more active component of picrotoxin, was used, obtained from Sigma), 50 mM L-glutamate or 300 mM sodium iodide. In an effort to minimize occupancy of chloride in the iodide-soaked crystals, crystals were transferred serially into three replicate cryoprotectant solutions lacking chloride before flash freezing. Nonetheless, because the iodide soaks were only 1–2 min, some chloride may have been carried over from crystallization. Electron density maps derived from these crystals showed no significant change in the strength of the electron density feature in the pore where we have modelled a chloride ion. We also soaked crystals in an analogous manner in bromide-containing cryo-solutions but were not able to observe significant peaks in the resultant anomalous difference electron density maps.

Data collection. Diffraction data were collected using synchrotron radiation at the Advanced Photon Source (Argonne National Laboratory, beamline 24-ID-C) with a mini-Kappa goniometer and in-house crystal alignment strategy software. The best-ordered crystals have a diffraction limit of 3.26 Å, a mosaic spread of 0.2–0.5°, and they are of the space group P4₃2₁2 with one GluCl_{cryst}–Fab complex

per asymmetric unit. The unit cell dimensions are $a = b = 155$ Å, $c = 575$ Å, $\alpha = \beta = \gamma = 90^\circ$, resulting in a Matthews' coefficient (V_M) of 4.0 Å³ Da^{−1} (ref. 55). Diffraction data were indexed, integrated and scaled using HKL2000 (ref. 56) or Xia2 (refs 57–62) software.

Structure determination. The structure was solved by molecular replacement using Phaser⁶³; the search probe was a pentameric homology model of GluCl_{cryst} made from GLIC (PDB code 3EHZ)¹⁰, using Swiss-Model⁶⁴. After an initial solution was found, phases were improved by solvent flattening⁶⁵ and electron density for Fab molecules bound at each of the five subunit interfaces of GluCl_{cryst} became plainly visible. A Fab homology model was made, using PDB 1NGQ for the light chain and 1F3D for the heavy chain, and Coot⁶⁶ to overlay the two modelled chains to make a single Fab molecule. Fab CDR loops were truncated and the model was used for molecular replacement using the GluCl_{cryst} solution as a starting point. In this manner, a single Fab was placed, and by copying the remaining Fab molecules around the fivefold non-crystallographic symmetry (NCS) axis, approximate positioning of all Fab molecules was accomplished. Electron density for Fab constant domain regions was poor after NCS averaging, and from non-averaged maps it was clear that the Fab constant domains did not obey five-fold symmetry. A starting model that included GluCl_{cryst} and five Fab variable domains was used for automated building with Buccaneer⁶⁷. Electron density maps were then good enough to position ivermectin molecules in the transmembrane domain loci, and to begin manual building of the Fab constant domains. Ivermectin stereochemistry, determined previously, is modelled as such⁶⁸. Numbering of ivermectin atoms in the figures is as defined in the PDB files; Supplementary Table 2 relates this numbering to that from the small molecule structure.

Iterative refinement of the model against the X-ray data using Phenix⁶⁹, manual adjustment in Coot into simulated annealing composite omit electron density maps⁶⁹ or real-space averaged maps⁷⁰, and structure quality analysis using Molprobity⁷¹ were carried out until satisfactory model statistics were obtained. Three groups of fivefold NCS restraints were present during refinement: five subunits of GluCl_{cryst}, five heavy chain Fab variable domains (residues 1–120), and the five light chain Fab variable domains (residues 1–108); the root mean squared deviation (r.m.s.d.) values between the chains within each of these three groups were 0.017, 0.014 and 0.015 Å, respectively. Isotropic B factors and TLS parameters were also refined; the 15 TLS groups comprised five GluCl_{cryst} subunits, five Fab variable domains, and five Fab constant domains. The final models contain the GluCl_{cryst} pentamer from residues 1–339 or 340, five ivermectin molecules, a single N-linked carbohydrate at N185 in three of the five subunits, five Fab molecules (1–221 for heavy chains, 1–210 for light chains), and several lipid and detergent molecules. Some portions of the Fab constant domains lacked electron density in composite omit maps and hence were omitted from the final model. The iodide-bound structure is of very low resolution and not completely refined: several anomalous difference electron density peaks in the extracellular domain were not modelled with iodide atoms.

Sequence alignments were made using PROMALS3D (ref. 72) and ClustalW⁷³. Isoelectric surface calculations were made using the APBS⁷⁴ add-on in PyMOL⁷⁵. Pore dimensions were analysed using HOLE software⁷⁶.

Electrophysiology. RNAs encoding GluCl proteins were transcribed from pGEM-HE⁷⁷ plasmids using the mMessage mMachine T7 Ultra kit (Ambion). Defolliculated stage V–VI *Xenopus* oocytes were provided by D. C. Dawson and C. Alexander, prepared as previously described⁷⁸. Oocytes were injected with 25 ng of GluCl RNAs, and current recordings were made 3–5 days afterwards. Frog saline (FS) recording solution contained (in mM): 96 NaCl, 2 KCl, 1 MgCl₂, 1.8 CaCl₂, 5 HEPES pH 7.5. Recording solution for iodide permeability experiments was FS but with NaI in place of NaCl. All ligands were made up in FS from stock solutions in water, except: picrotoxin, 1 M stock in DMSO; ivermectin, 5 mM stock in DMSO. Recording electrode pipettes (0.7–2 M Ω) were cushioned with 0.8% LMP agarose in 3 M KCl and backfilled with 3 M KCl. Oocytes were voltage-clamped at −80 mV except in experiments to determine the reversal potential, which used 40 ms voltage steps. Analogue data were filtered at 50 Hz and digitized at ≥ 1 kHz. The Axoclamp 2B amplifier (Axon Instruments) and pClamp 10 software (Molecular Devices) were used for data acquisition. In un-injected oocytes, no significant responses to test solutions were observed (Supplementary Fig. 13).

Radioligand binding experiments. Experiments to test binding of [³H]-L-glutamate to GluCl_{cryst} and competition of the radioligand with other compounds were done using purified Nano15-tagged⁷⁹ GluCl_{cryst} (N-terminal tag) and streptavidin–Ysi scintillation proximity assay beads (SPA; GE Healthcare Life Sciences). The concentration of binding sites was fixed at 100 nM after a preliminary experiment to determine optimal GluCl_{cryst} concentration (Supplementary Fig. 12). Other binding assay components were: 50 mM Tris pH 7.4, 150 mM NaCl, 1 mM C₁₂M, 1 mg ml^{−1} SPA beads, and 1 μ M ivermectin. Saturation

binding of [^3H]-L-glutamate in the presence and absence of Fab was performed with a 1:30 dilution of specific activity of the radiolabel with [^1H]-L-glutamate, and a slight molar excess of Fab to GluCl_{cryst} subunits as verified by FSEC experiments. Measurement of background signal in saturation binding experiments was complicated by, we believe, significant binding of [^3H]-L-glutamate directly to SPA beads and lack of a chemically distinct competitor for the neurotransmitter-binding site. Neither high concentrations of [^1H]-L-glutamate or absence of protein were able to fully account for this apparently non-specific signal. To address the background component that was not accurately measured experimentally, we combined subtraction of a background signal measured in the absence of GluCl_{cryst} with a linear component still present in the binding data (calculated using the total binding function in the fitting software). In saturation binding experiments in the presence of Fab (Supplementary Fig. 12), data were better fit after removing background signal measured in the presence of 10 mM [^1H]-L-glutamate combined with the calculated linear component. In competition binding experiments to determine IC₅₀ values, [^3H]-L-glutamate was 1 μM using a 1:10 dilution of specific activity of the radiolabel with cold glutamate. In all [^3H]-L-glutamate and electrophysiological dose-response experiments, data were fit with GraphPad Prism software.

51. Gensler, S. *et al.* Assembly and clustering of acetylcholine receptors containing GFP-tagged ϵ or γ subunits: selective targeting to the neuromuscular junction *in vivo*. *Eur. J. Biochem.* **268**, 2209–2217 (2001).
52. Li, P., Slimko, E. M. & Lester, H. A. Selective elimination of glutamate activation and introduction of fluorescent proteins into a *Caenorhabditis elegans* chloride channel. *FEBS Lett.* **528**, 77–82 (2002).
53. Slimko, E. M. & Lester, H. A. Codon optimization of *Caenorhabditis elegans* GluCl ion channel genes for mammalian cells dramatically improves expression levels. *J. Neurosci. Methods* **124**, 75–81 (2003).
54. Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1988).
55. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497 (1968).
56. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
57. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
58. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
59. Leslie, A. G. W. Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography, No. 26 (1992).
60. Leslie, A. G. The integration of macromolecular diffraction data. *Acta Crystallogr. D* **62**, 48–57 (2006).
61. Sauter, N. K., Grosse-Kunstleve, R. W. & Adams, P. D. Robust indexing for automatic data collection. *J. Appl. Cryst.* **37**, 399–409 (2004).
62. Zhang, C. Y., Sauter, N. K., van den Bedem, H., Snell, G. & Deacon, A. M. Automated diffraction image analysis and spot searching for high-throughput crystal screening. *J. Appl. Cryst.* **39**, 112–119 (2006).
63. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
64. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201 (2006).
65. Cowtan, K. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography **31**, 34–38 (1994).
66. Emsley, P. & Cowtan, P. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
67. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62**, 1002–1011 (2006).
68. Springer, J. P., Arison, B. H., Hirshfield, J. M. & Hoogsteen, K. The absolute stereochemistry and conformation of avermectin B_{2a} aglycon and avermectin B_{1a}. *J. Am. Chem. Soc.* **103**, 4221–4224 (1981).
69. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
70. Collaborative Computational Project 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–776 (1994).
71. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
72. Pei, J., Kim, B. H. & Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300 (2008).
73. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
74. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
75. DeLano, W. L. *The PyMOL Molecular Graphics System* (<http://www.pymol.org>) (DeLano Scientific, 2002).
76. Smart, O. S., Goodfellow, J. M. & Wallace, B. A. The pore dimensions of gramicidin A. *Biophys. J.* **65**, 2455–2460 (1993).
77. Liman, E. R., Tytgat, J. & Hess, P. Subunit stoichiometry of a mammalian K⁺ channel determined by construction of multimeric cDNAs. *Neuron* **9**, 861–871 (1992).
78. Alexander, C. *et al.* Cystic fibrosis transmembrane conductance regulator: using differential reactivity toward channel-permeant and channel-impermeant thiol-reactive probes to test a molecular model for the pore. *Biochemistry* **48**, 10078–10088 (2009).
79. Lamla, T. & Erdmann, V. A. The Nano-tag, a streptavidin-binding peptide for the purification and detection of recombinant proteins. *Protein Expr. Purif.* **33**, 39–47 (2004).

The thermodynamic meaning of negative entropy

Lidia del Rio^{1*}, Johan Åberg^{1*}, Renato Renner¹, Oscar Dahlsten^{1,2,3} & Vlatko Vedral^{2,3,4}

The heat generated by computations is not only an obstacle to circuit miniaturization but also a fundamental aspect of the relationship between information theory and thermodynamics. In principle, reversible operations may be performed at no energy cost; given that irreversible computations can always be decomposed into reversible operations followed by the erasure of data^{1,2}, the problem of calculating their energy cost is reduced to the study of erasure. Landauer's principle states that the erasure of data stored in a system has an inherent work cost and therefore dissipates heat^{3–8}. However, this consideration assumes that the information about the system to be erased is classical, and does not extend to the general case where an observer may have quantum information about the system to be erased, for instance by means of a quantum memory entangled with the system. Here we show that the standard formulation and implications of Landauer's principle are no longer valid in the presence of quantum information. Our main result is that the work cost of erasure is determined by the entropy of the system, conditioned on the quantum information an observer has about it. In other words, the more an observer knows about the system, the less it costs to erase it. This result gives a direct thermodynamic significance to conditional entropies, originally introduced in information theory. Furthermore, it provides new bounds on the heat generation of computations: because conditional entropies can become negative in the quantum case, an observer who is strongly correlated with a system may gain work while erasing it, thereby cooling the environment.

'Erasure' of a system is defined as taking it to a pre-defined pure state, $|0\rangle$ (a familiar example is disk formatting, where a sequence of zero bits is written onto the disk). Landauer's principle asserts that the energy dissipated to erase a system, S , using an optimal process in an environment of temperature T is given by

$$W(S) = H(S)kT\ln(2) \quad (1)$$

where k is the Boltzmann constant^{3,8–13}. Here the von Neumann entropy, $H(S) = -\text{Tr}[\rho\log_2(\rho)]$, quantifies the uncertainty about system S , whose state is described by ρ .

Different observers may have different knowledge about the same system. For instance, an observer, Alice, can prepare n quantum bits (qubits; for example n spin-1/2 particles) in a pure state of her choice, keeping a record of that state. However, from the point of view of another observer, Bob, who does not have access to Alice's record, the state of the system is completely unknown: he describes it as a fully mixed state, of maximal entropy. Hence, rather than $W(S)$ being defined as the 'cost of erasure', it may be described as the 'cost of erasure for observer C ', which we denote by $W(S|C)$. This leads to the following reformulation of equation (1), where $H(S|C)$ denotes the uncertainty that observer C has about system S :

$$W(S|C) = H(S|C)kT\ln(2) \quad (2)$$

Typically, the observer C is assumed to be classical. More explicitly, classical observers represent their information about S in a memory that consists of classical bits (as in the case of Alice and Bob above).

Our contribution is to go beyond this classical picture and consider observers who may have access to information that is represented as the state of a quantum system: a quantum memory. As an example, we could imagine a third observer, Quasimodo, who has a memory that includes n qubits, each maximally entangled with a particle of S .

Our main result characterizes the work, $W(S|Q)$, that an observer with access to a quantum memory, Q , needs to perform to erase system S . For simplicity, we formulate it here for a special case, which could be referred to as a 'thermodynamic limit' of erasure, where the observer erases many identical copies of S jointly. In this case, we show that there exists an erasure process whose work cost does not exceed

$$W(S|Q) = H(S|Q)kT\ln(2) \quad (3)$$

per copy of S . Here $H(S|Q)$ is the conditional von Neumann entropy, $H(S|Q) = H(SQ) - H(Q)$. We show that this work cost is optimal, under the assumption that Landauer's principle holds for a classical observer (Methods Summary). In its general form, our main result (Theorem 1 in Methods Summary) includes the more natural case where a single set of data, rather than a collection of identical copies, is to be erased.

Crucially, we require that the information stored in Q be preserved by the erasure process—a non-trivial condition, given that accessing a quantum memory can in principle change it. This information-preservation property is particularly important if we consider the erasure process to be part of a larger procedure (see, for instance, Fig. 1). For example, suppose that we erase system S and later would like to erase another system, Z . If the erasure of S removed the information about Z , erasing Z could become unnecessarily costly. More generally, we can consider a reference system, R , that models all other systems about which the memory can have information (technically, R is a purification of Q and S). The information-preservation condition can then be formulated as the requirement that the joint state of the memory and the reference system, ρ_{QR} , be preserved by the erasure process (see also Fig. 2).

The generalization of erasure to the quantum case exposes features not present in a classical setting. In particular, equation (3) implies that the work required for erasure may be negative for an observer with a quantum memory: the process results in a net gain of work. This happens because the uncertainty $H(S|Q)$ can become negative for quantum observers. For instance, Quasimodo's conditional von Neumann entropy is $H(S|Q) = 0 - n$ (because the joint state of S and Q is pure and the reduced state of the memory, Q , is fully mixed). Our result provides a thermodynamic operational meaning for negative conditional entropies, which until now only had information-theoretical interpretations; for instance, they measure the entanglement needed to send a state to a receiver with side information¹⁴ (state merging), and quantify 'violations' of Heisenberg's uncertainty principle¹⁵.

The proof of equation (3) uses a probabilistic method to find appropriate erasure procedures. In the simple examples of our three observers, Alice, Bob and Quasimodo, we can describe them explicitly. Alice, who holds a classical description of the pure state of S , has no uncertainty about the system: $H(S|A) = 0$. As expected, she does not need to

¹Institute for Theoretical Physics, ETH Zurich, 8093 Zurich, Switzerland. ²Clarendon Laboratory, University of Oxford, Oxford OX1 3PU, UK. ³Centre for Quantum Technologies, National University of Singapore, 117543 Singapore. ⁴Department of Physics, National University of Singapore, 117542 Singapore.

*These authors contributed equally to this work.

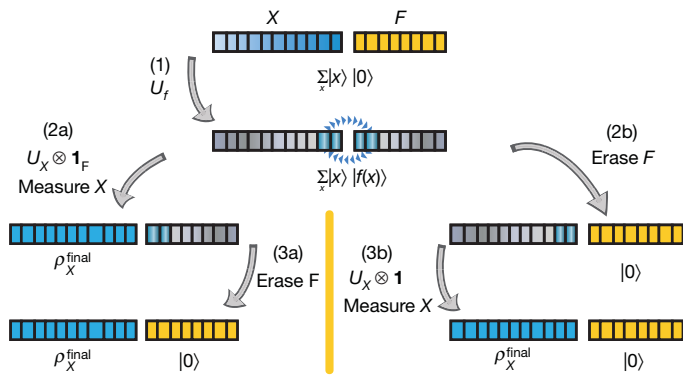


Figure 1 | Erasure in quantum computation. As an example, we consider the period-finding algorithm for a function f , used in the quantum part of Shor's factoring algorithm. (1) Initially the algorithm evaluates f on a superposition of all inputs, x , creating an entangled state between two registers, X and F (this is done with a unitary transformation, U_f , on the two registers). Given statistical knowledge about the properties of f , it is possible to find highly correlated subsystems of X and F (in blue). (2a) The second part of the algorithm consists only of local operations on X (a unitary transformation, U_X , followed by a measurement; the final state of X is ρ_X^{final}). (3a) Usually F is erased at the end of the algorithm, when correlations between X and F have been destroyed. (2b) Instead, it is possible to erase F while it is still partly entangled with X , using the correlations (in blue) to decrease the work cost of the erasure. If the erasure can be performed in a computationally efficient manner, it may be incorporated in the algorithm. (3b) The information-preservation condition ensures that the rest of the algorithm is not affected by the early erasure. For a concrete example and further discussion, see Supplementary Information, section V.

do any work to erase S : she may consult her record to check the state of S and change it to $|0\rangle$ with a reversible transformation (Fig. 3).

Bob, however, has no access to Alice's record and, so, has maximal uncertainty: $H(S|B) = n$. He can perform a simple erasure process^{10,11} that lets the system interact with a heat bath at temperature T (Fig. 4). The work cost of this process is $nkT\ln(2)$.

Turning now to Quasimodo, recall that his memory contains n qubits maximally entangled with S . We call this part of his memory Q_1 and denote the entangled state $|SQ_1\rangle$. In addition, the rest of his memory, Q_2 , is correlated with a reference system, R , in state $|Q_2R\rangle$. To erase S , Quasimodo combines two elementary procedures: the erasure process used by Bob, and its reverse, a 'work-extraction' process, whereby he transforms an initially pure state into a maximally mixed state, gaining energy $kT\ln(2)$ per qubit (Fig. 4). He starts by performing the elementary work extraction on the $2n$ -qubit state $|SQ_1\rangle$ to gain

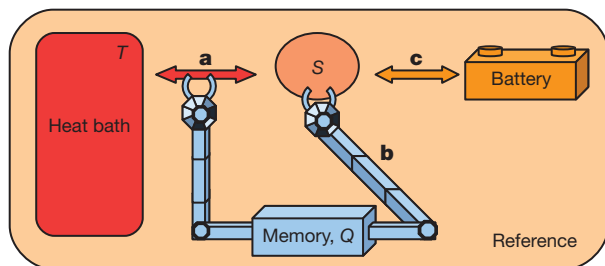


Figure 2 | Erasure setting. An observer, here represented by a machine with a quantum memory, Q , erases a system, S . All memory contents about a reference system must be preserved. We assume that the initial Hamiltonian of S and Q is fully degenerate (for example like that of paramagnets in a zero magnetic field). We use a simple model for erasure, with the following options. **a**, The observer may couple S to a heat bath at temperature T ; the bath thermalizes S , leaving it in a Gibbs state. **b**, The observer may manipulate S and Q , by (1) applying unitary operations to those systems, and (2) raising or lowering any energy levels of their Hamiltonian (for example by tuning a magnetic field). By raising or lowering an occupied level by ΔE , the observer uses or, respectively, gains energy ΔE ; empty levels can be changed at no energy cost. **c**, The observer may store energy in and withdraw it from a battery.

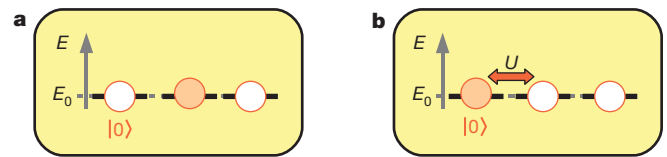


Figure 3 | Erasure of a pure state. The circles represent the energy eigenstates of system S , and a filled circle means that the system can be found in that state with certainty. **a**, Alice knows that the system is in a particular pure state. **b**, She performs a unitary transformation, U , that swaps that state with $|0\rangle$. If the Hamiltonian of S is degenerate, this operation has no energy cost.

energy $2nkT\ln(2)$, leaving S and Q_1 maximally mixed. Then he performs the elementary erasure process on S , which costs him $nkT\ln(2)$ in energy. In total, his energy gain is $nkT\ln(2)$, matching the prediction of equation (3). The final state of Quasimodo's memory and the reference is $\rho_{QR} = 2^{-n} \mathbf{1}_{Q_1} \otimes |Q_2R\rangle\langle Q_2R|$, where $2^{-n} \mathbf{1}_{Q_1}$ is the fully mixed state on Q_1 . This is precisely the initial state of QR , because the original state of Q_1S had a fully mixed marginal in Q_1 —the information-preservation condition is therefore satisfied.

In general, correlations between S and the memory are not as neat as for Quasimodo. Nevertheless, this special example contains the essence of the general case. Using decoupling results¹⁶, we show that, independently of the exact form of the correlations between S and Q , it is possible to find a subsystem of S and Q that is (approximately) in a pure state. This pure subsystem furthermore has a special structure that allows us to extract work from it, thus replacing it with a maximally mixed state, without changing the state of the memory and the reference (Fig. 5). The size of the pure state found and, therefore, the work gained depends on the entropy of S conditioned on the information held by the quantum memory (see Supplementary Information, section I, for a formal proof). In its general single-shot form, the erasure procedure we introduce has a failure probability that can be made arbitrarily small, at the cost of increasing the work performed by an additive term (Theorem 1 in Methods Summary; non-deterministic work extraction has been discussed before for classical observers¹⁷).

The erasure processes contemplated in this work require considerable control over the quantum systems involved, and it may not be clear why such idealizations are interesting. As an analogy we can think of the Carnot cycle. Although the ideal performance of a Carnot engine may be an unattainable limit in practice, it nevertheless provides the theoretical foundation for the behaviour of real heat engines. In a similar spirit, our investigation bounds the ideal minimum work cost for the implementation of irreversible processes.

Such arguments are particularly relevant to the study of heat generation in computation, one of the major obstacles to the miniaturization of circuitry. Computation can in principle be made reversible,

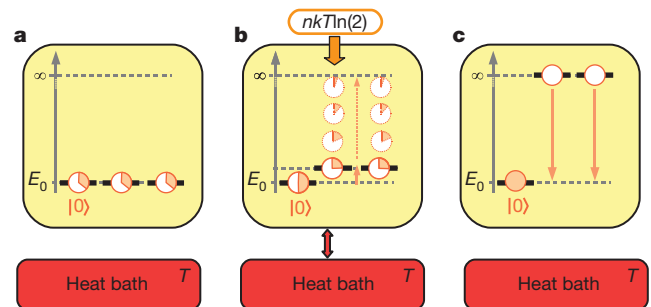


Figure 4 | Erasure of a fully mixed state and work extraction. **a**, Initially, S is in a fully mixed state. **b**, Keeping one energy level untouched, Bob raises all other levels in small steps, letting S thermalize after each step. As these levels are raised it is more likely to find S in the untouched level. **c**, Bob decouples S from the bath and lowers all states back to the original level, in one single step. The work cost of the whole process in the quasistatic limit is $kT\ln(2)$ per qubit. By running this process in reverse, we obtain a work-extraction procedure, which given an initial pure state yields an energy gain of $kT\ln(2)$ per qubit, at the cost of leaving the state maximally mixed.

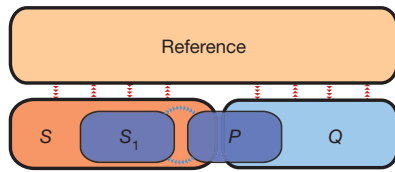


Figure 5 | General erasure procedure. The erasure proceeds in three steps. First we find a subsystem, S_1 , that is decoupled from the reference; the reduced state of S_1 is approximately fully mixed¹⁶. The size of S_1 is limited by the correlations between S and the reference. These are weak if S is highly correlated with Q (because the global state is pure²⁸), such that $\log_2(|S_1|) \approx [\log_2(|S|) - H(S|Q)]/2$. Because S_1 is decoupled from the reference, it is purified by a subsystem, P , of $S \otimes Q$. The state of $S_1 \otimes P$ is maximally entangled. Then we extract work from the pure state of $S_1 \otimes P$, gaining energy $2\log_2(|S_1|)kT\ln(2)$. The reduced state of P , originally fully mixed, is preserved by this process. Finally we erase S , performing work $\log_2(|S|)kT\ln(2)$. The total work cost of the process is approximately $H(S|Q)kT\ln(2)$. See Supplementary Information for details.

but at the expense of keeping extra data in a memory^{1,2}. A part of that memory may then be erased, keeping the rest intact. Naively, the work cost of that operation, and the corresponding heat generation, would be given solely by the entropy of the part to be erased. However, our analysis shows that erasure can be optimized if information stored in other parts of the memory is used (Fig. 1).

The correspondence between conditional entropy and work that we found can also be used to quantify quantum correlations¹⁸. More precisely, because an energy gain in erasure relies on entanglement between the system and the memory, an erasure process with negative work cost can serve as an entanglement witness¹⁹. Similarly, our work is related to discord^{20–23}. Our results suggest that discord can quantify the difference between the respective work costs of erasure using quantum and classical memories²⁴. Because our relation is valid for a single instance of erasure, it may be used to obtain a single-shot generalization of discord.

In this work, we have used information-theoretical tools, such as decoupling^{14,16,25}, to prove a physical result. It is also possible to translate thermodynamic statements into information-theoretical ones. For instance, the work required to erase a system S cannot be reduced by locally processing information about S (see Supplementary Information, Lemma I.6). Using our results, we can infer that the conditional entropy $H(S|Q)$ cannot decrease under local operations on Q , which is a fundamental result in information theory known as the data processing inequality. In general, we expect our results to strengthen the link between information theory and statistical physics.

METHODS SUMMARY

Here we characterize a single instance of erasure. The work cost of erasure, $W(S|Q)$, is a random variable and may fluctuate. Our main result is an upper bound for $W(S|Q)$ that is violated only with a very small probability. This we state as Theorem 1: there exists a process to erase a system S , conditioned on memory Q , at temperature T , whose work cost satisfies

$$W(S|Q) \leq [H_{\max}^{\varepsilon}(S|Q) + A]kT\ln(2)$$

except with probability less than $\delta = \sqrt{2^{-A/2} + 12\varepsilon}$ for all $A \geq 0$ and all $\varepsilon \geq 0$.

The quantity $H_{\max}^{\varepsilon}(S|Q)$ denotes the smooth max-entropy of system S conditioned on the quantum memory, Q , and is a single-shot generalization of the von Neumann entropy²⁶ (Supplementary Information, section II).

We can fix δ and then adjust ε to minimize the work cost. For instance, if we allow a probability of failure of 3%, we pay an extra price of $A \approx 20$ in the total work cost, independently of the size of S .

The proof of Theorem 1 is sketched in Fig. 5 and can be found in Supplementary Information, section I.

To understand the exact meaning of equation (3), we consider a ‘thermodynamic limit’ of erasure, where a large collection of independent and identical copies of the system is erased. We define the ‘work cost rate’ as the average $\bar{w}(S|Q) = (1/n)W(S^{\otimes n}|Q^{\otimes n})$ for an optimal erasure process, in the limit of large n (Supplementary Information, Definition I.3). A result from information theory, the quantum asymptotic equipartition property²⁷, asserts that the smooth

max-entropy converges to the von Neumann entropy for many identical copies of the state (Supplementary Information, section II.B). Combining Theorem 1 with the asymptotic equipartition property, we find that there exists an erasure process such that $\bar{w}(S|Q) \leq H(S|Q)kT\ln(2)$; hence, the average work cost never exceeds $W(S|Q) = H(S|Q)kT\ln(2)$. Furthermore, if we assume that Landauer’s principle holds for a classical observer (equation (2)), we can show that the quantum bound is tight in this limit (Supplementary Information, Lemma I.5).

Received 24 November 2010; accepted 18 April 2011.

- Bennett, C. H. Logical reversibility of computation. *IBM J. Res. Develop.* **17**, 525–532 (1973).
- Fredkin, E. & Toffoli, T. Conservative logic. *Int. J. Theor. Phys.* **21**, 219–253 (1982).
- Landauer, R. Dissipation and heat generation in the computing process. *IBM J. Res. Develop.* **5**, 148–156 (1961).
- Bennett, C. H. The thermodynamics of computation — a review. *Int. J. Theor. Phys.* **21**, 905–940 (1982).
- Leff, H. S. & Rex, A. F. *Maxwell’s Demon: Entropy, Information, Computing* (Taylor and Francis, 1990).
- Leff, H. S. & Rex, A. F. *Maxwell’s Demon 2: Entropy, Classical and Quantum Information, Computing* (Taylor and Francis, 2002).
- Plenio, M. B. & Vitelli, V. The physics of forgetting: Landauer’s erasure principle and information theory. *Contemp. Phys.* **42**, 25–60 (2001).
- Maruyama, K., Nori, F. & Vedral, V. The physics of Maxwell’s demon and information. *Rev. Mod. Phys.* **81**, 1–23 (2009).
- Bennett, C. H. Notes on Landauer’s principle, reversible computation and Maxwell’s demon. *Stud. Hist. Phil. Mod. Phys.* **34**, 501–510 (2003).
- Alicki, R., Horodecki, M., Horodecki, P. & Horodecki, R. Thermodynamics of quantum information systems — Hamiltonian description. *Open Syst. Inf. Dyn.* **11**, 205–217 (2004).
- Piechocinska, B. Information erasure. *Phys. Rev. A* **61**, 062314 (2000).
- Shizume, K. Heat generation required by information erasure. *Phys. Rev. E* **52**, 3495–3499 (1995).
- Schulman, L. J. & Vazirani, U. V. in *Proc. 31st Annu. ACM Symp. Theory Comput.* (eds Vitter, J. S., Larmore, L. & Leighton, T.) 322–329 (Association for Computing Machinery, 1999).
- Horodecki, M., Oppenheim, J. & Winter, A. Partial quantum information. *Nature* **436**, 673–676 (2005).
- Berta, M., Christandl, M., Colbeck, R., Renes, J. M. & Renner, R. The uncertainty principle in the presence of quantum memory. *Nature Phys.* **6**, 659–662 (2010).
- Dupuis, F. *The Decoupling Approach to Quantum Information Theory*. PhD thesis, Univ. de Montréal (2009).
- Dahlsten, O., Renner, R., Rieper, E. & Vedral, V. Inadequacy of von Neumann entropy for characterising extractable work. *N. J. Phys.* **13**, 053015 (2011).
- Oppenheim, J., Horodecki, M., Horodecki, P. & Horodecki, R. Thermodynamical approach to quantifying quantum correlations. *Phys. Rev. Lett.* **89**, 180402 (2002).
- Horodecki, M., Horodecki, P. & Horodecki, R. Separability of mixed states: necessary and sufficient conditions. *Phys. Lett. A* **223**, 1–8 (1996).
- Zurek, W. H. Einselection and decoherence from an information theory perspective. *Ann. Phys.* **9**, 853–862 (2000).
- Ollivier, H. & Zurek, W. H. Quantum discord: a measure of the quantumness of correlations. *Phys. Rev. Lett.* **88**, 017901 (2000).
- Cavalcanti, D., Aolita, L. & Boixo, S. Modi, K. Piani, M. & Winter, A. Operational interpretations of quantum discord. *Phys. Rev. A* **83**, 032324 (2011).
- Modi, K., Paterek, T., Son, W., Vedral, V. & Williamson, M. Unified view of quantum and classical correlations. *Phys. Rev. Lett.* **104**, 080501 (2010).
- Zurek, W. H. Quantum discord and Maxwell’s demons. *Phys. Rev. A* **67**, 012320 (2003).
- Hayden, P. & Preskill, J. Black holes as mirrors: quantum information in random subsystems. *J. High Energy Phys.* **09**, 120 (2007).
- Renner, R. *Security of Quantum Key Distribution*. PhD thesis, ETH Zurich (2005).
- Tomamichel, M., Colbeck, R. & Renner, R. A fully quantum asymptotic equipartition property. *IEEE Trans. Inf. Theory* **55**, 5840–5847 (2009).
- Tomamichel, M., Colbeck, R. & Renner, R. Duality between smooth min- and max-entropies. *IEEE Trans. Inf. Theory* **56**, 4674–4681 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Colbeck for discussions. We acknowledge support from the Swiss National Science Foundation (L.d.R., J.A., R.R. and O.D.; grant no. 200021-119868 and the NCCR QSIT), the Portuguese Fundação para a Ciência e Tecnologia (L.d.R.; grant no. SFRH/BD/43263/2008), the European Research Council (R.R.; grant no. 258932) and Singapore’s National Research Foundation and Ministry of Education (V.V.).

Author Contributions The main ideas were developed by all authors. L.d.R., J.A. and R.R. formulated and proved the main technical claims. L.d.R. and J.A. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.d.R. (delrio@phys.ethz.ch).

A graphene-based broadband optical modulator

Ming Liu^{1*}, Xiaobo Yin^{1*}, Erick Ulin-Avila¹, Baisong Geng², Thomas Zentgraf¹, Long Ju², Feng Wang^{2,3} & Xiang Zhang^{1,3}

Integrated optical modulators with high modulation speed, small footprint and large optical bandwidth are poised to be the enabling devices for on-chip optical interconnects^{1,2}. Semiconductor modulators have therefore been heavily researched over the past few years. However, the device footprint of silicon-based modulators is of the order of millimetres, owing to its weak electro-optical properties³. Germanium and compound semiconductors, on the other hand, face the major challenge of integration with existing silicon electronics and photonics platforms^{4–6}. Integrating silicon modulators with high-quality-factor optical resonators increases the modulation strength, but these devices suffer from intrinsic narrow bandwidth and require sophisticated optical design; they also have stringent fabrication requirements and limited temperature tolerances⁷. Finding a complementary metal-oxide-semiconductor (CMOS)-compatible material with adequate modulation speed and strength has therefore become a task of not only scientific interest, but also industrial importance. Here we experimentally demonstrate a broadband, high-speed, waveguide-integrated electroabsorption modulator based on monolayer graphene. By electrically tuning the Fermi level of the graphene sheet, we demonstrate modulation of the guided light at frequencies over 1 GHz, together with a broad operation spectrum that ranges from 1.35 to 1.6 μm under ambient conditions. The high modulation efficiency of graphene results in an active device area of merely 25 μm^2 , which is among the smallest to date. This graphene-based optical modulation mechanism, with combined advantages of compact footprint, low operation voltage and ultrafast modulation speed across a broad range of wavelengths, can enable novel architectures for on-chip optical communications.

Graphene, a single sheet of carbon atoms in a hexagonal lattice, has attracted great interest because of its exceptional electrical and optical properties^{8–13}. Graphene couples strongly to light, which enables observation of monolayer graphene under an optical microscope with the naked eye. A pristine graphene monolayer has a constant absorption of $\pi\alpha = 2.3\%$ across the infrared and visible range, where α is the fine-structure constant ($e^2/\hbar c$, where e is the electronic charge, \hbar is Planck's constant divided by 2π , and c the velocity of light). Moreover, the broad optical absorption in graphene can be controlled through electrical gating: by shifting the electronic Fermi level, one can controllably change graphene's optical transitions^{14,15}. The strong electroabsorption effect, which has not yet been observed in bulk materials, originates from graphene's unique electronic structure and its two-dimensional character. It implies that graphene has the potential to be used as the active medium in an optical electroabsorption modulator. However, one of the challenges involved in a direct graphene modulator is the limited absorption of a monolayer. This can be overcome by integrating graphene with an optical waveguide, which greatly increases the interaction length through the coupling between the evanescent waves and graphene.

A graphene-based waveguide-integrated electroabsorption modulator has several distinctive advantages. (1) Strong light–graphene interaction. In comparison to compound semiconductors, such as

those exhibiting the quantum-well with quantum-confined Stark effect (QCSE)⁶, a monolayer of graphene possesses a much stronger interband optical transition, which finds applications in novel optoelectronic devices such as photodetectors^{16,17}. (2) Broadband operation. As the high frequency dynamic conductivity for Dirac fermions is constant, the optical absorption of graphene is independent of wavelength, covering all telecommunications bandwidth and also the mid- and far-infrared^{18,19}. (3) High-speed operation. With a carrier mobility exceeding 200,000 $\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$ at room temperature^{20,21} (this is among the highest known), the Fermi level and hence the optical absorptions of graphene can be rapidly modulated through the band-filling effect. In addition, speed limiting processes in graphene (such as photocarrier generation and relaxation) operate on the timescale of picoseconds²², which implies that graphene-based electronics may have the potential to operate at 500 GHz, depending on the carrier density and graphene quality. (4) Compatibility with CMOS processing. The athermal optoelectronic properties of graphene and its CMOS-compatible integration processes at wafer scale^{23–25} make it a promising candidate for post-CMOS electronics, particularly for high frequency applications. With all these advantages, monolithic integration of a graphene electroabsorption modulator could open new routes to integrated photonics, with a compact footprint, low voltage operation and ultrafast modulation across a broad range of wavelengths.

Here we report the first waveguide-integrated graphene-based electroabsorption modulator, in which modulation is achieved by actively tuning the Fermi level of a monolayer graphene sheet. The gigahertz graphene modulator demonstrates a strong electroabsorption modulation of 0.1 $\text{dB} \mu\text{m}^{-1}$ and operates over a broad range of wavelength, from 1.35 μm to 1.6 μm , under ambient conditions.

The structure of the electroabsorption modulator is schematically illustrated in Fig. 1. A 50-nm-thick Si layer was used to connect the 250-nm-thick Si bus waveguide and one of the gold electrodes. Both the silicon layer and the waveguide were shallowly doped with boron to reduce the sheet resistance. A spacer of 7-nm-thick Al_2O_3 was then uniformly deposited on the surface of the waveguide by atom layer deposition. A graphene sheet grown by chemical vapour deposition^{23,24,26} (CVD) was then mechanically transferred onto the Si waveguide. To further reduce the access resistance of the device, the counter electrode was extended towards the bus waveguide by depositing a platinum (10 nm) film on top of the graphene layer. The minimum distance between platinum electrode and waveguide was controlled at 500 nm, so that the optical modes of the waveguide remained undisturbed by the platinum contact. The excess graphene was removed by oxygen plasma, leaving only the regions on top of the waveguide and between the waveguide and the platinum electrode.

The cross-sectional view of the device structure and the optical field distribution of the guided mode are shown in Fig. 1b. The thin silicon layer and the platinum electrode adjacent to the waveguide have negligible effect on the mode profile. To further improve the electroabsorption modulation efficiency, the silicon waveguide was designed to have the electric field maximized at its top and bottom surfaces, so that the interband transitions in graphene are also maximized (Fig. 1b). As

¹NSF Nano-scale Science and Engineering Center (NSEC), 3112 Etcheverry Hall, University of California at Berkeley, Berkeley, California 94720, USA. ²Department of Physics, University of California at Berkeley, Berkeley, California 94720, USA. ³Materials Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA.

*These authors contributed equally to this work.

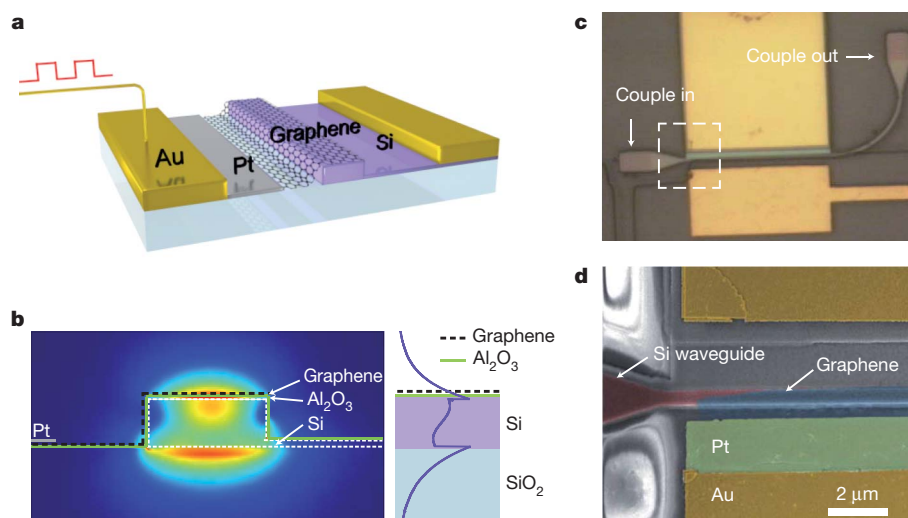


Figure 1 | A graphene-based waveguide-integrated optical modulator. **a**, Three-dimensional schematic illustration of the device; a monolayer graphene sheet is on top of a silicon bus waveguide, separated from it by a 7-nm-thick Al_2O_3 layer (not shown). The silicon waveguide is doped, and connected to the electrode (right, shown gold) through a thin layer of silicon defined by selective etching. **b**, Left, cross-section of the device, with an overlay of the optical mode plot, calculated by finite element simulation. The waveguide was carrying a single optical mode, and was designed so as to maximize the field at the interface between the waveguide and the graphene, to maximize the absorption efficiency. The thin silicon layer and the Pt electrode, which is 500 nm away from the waveguide, have negligible influence on the optical mode. Right, a cross-section through the centre of the waveguide; the purple

curve shows the magnitude of the electric field. The actual thicknesses of the graphene sheet and the Al_2O_3 in the simulation are 0.7 nm and 7 nm, respectively. **c**, Top-view optical microscope image of the waveguide. The Si waveguide was bent by 90° to change the polarization state between the input and the output light, to improve the signal-noise ratio. **d**, SEM image showing the boxed region in **c** at higher magnification, showing the detailed structure of the graphene modulator. False colours are used to highlight the Au electrode (yellow), the Pt electrode (green), the graphene sheet (blue) and the waveguide (red). The width of the Si waveguide is 600 nm, while the distance between the Pt electrode and the Si waveguide is 500 nm. The bright multi-ring region beside the Au electrodes is due to a charging effect on the SiO_2 layer in the SEM.

graphene only interacts with the tangential (in-plane) electric field of electromagnetic waves, the graphene modulator is polarization-sensitive, as are conventional semiconductor-based electro-optical modulators³.

Figure 1c shows a top-view optical microscope image of the device, and a close-up scanning electron microscopy image of the active region is given in Fig. 1d. The graphene sheet, highlighted in Fig. 1d by a false blue colour, covers only the waveguide region; this is to minimize the capacitance. The platinum electrode (green) is placed 500 nm away from the 600-nm-wide Si waveguide. Light was coupled in and out of the waveguide through tapered gratings, which contribute most to the overall loss of the system. The Si waveguide was bent 90° to change the polarization state between the input and the output light, to improve the signal-noise ratio.

Figure 2 displays the transmission of $1.53\text{ }\mu\text{m}$ photons through the waveguide at different drive voltages, V_D . At low drive voltage ($-1\text{ V} < V_D < 3.8\text{ V}$), the Fermi level $E_F(V_D)$ of graphene is close to the Dirac point ($E_F(V_D) < \hbar\nu_0/2$), and interband transitions occur when electrons are excited by the incoming photons ($\hbar\nu_0$). The optical absorption of graphene is determined by the position of the Fermi level. By applying a drive voltage between the graphene and the waveguide, we can tune the Fermi level of the graphene, and therefore modulate the total transmission. With the current waveguide design, the modulation depth is as high as $0.1\text{ dB }\mu\text{m}^{-1}$, resulting in a graphene electroabsorption modulator with a footprint of merely $25\text{ }\mu\text{m}^2$. At large negative V_D ($< -1\text{ V}$), the Fermi level is lowered below the transition threshold ($E_F(V_D) = \hbar\nu_0/2$) owing to positive charge accumulation. As a result, there are no electrons available for interband transitions, and hence the graphene appears transparent. On the other hand, at large positive V_D ($> 3.8\text{ V}$), all electron states are filled up, and no interband transitions are allowed. Ideally, there should be a sharp change in transmission at $E_F(V_D) = \hbar\nu_0/2$. In reality, this transition was broadened owing to defects in the graphene, and shifted to higher voltage owing to natural doping from the substrate²⁷. When the modulator is in operation (that is, when no interband absorption is allowed),

its insertion loss is negligible as the intraband absorption of graphene is extremely low at near-infrared wavelengths²⁸.

To measure the dynamic response of the graphene modulator, radio frequency signals generated by a network analyser were added on a static V_D and applied to the modulator. The same $1.53\text{-}\mu\text{m}$ laser was used to test the modulator, and the out-coupled light was sent to a high-speed photodetector. Shown in Fig. 3 are the V_D -dependent r.f. responses of the graphene modulator; gigahertz operation of the device

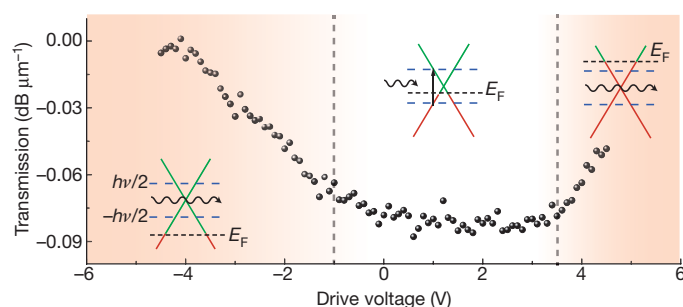


Figure 2 | Static electro-optical response of the device at different drive voltages. The main panel shows the modulation depth, normalized to the device length ($40\text{ }\mu\text{m}$), at different drive voltages (V_D). Three regions can be seen, and their band structures are shown as insets. In the middle region (with V_D in the range -1 V to 3.8 V), the Fermi level (E_F , black dashed line) is close to the Dirac point and the interband transition is allowed from electron-occupied regions (red lines) to unoccupied regions (green lines). Thus the graphene sheet is absorptive to incident photons ($\hbar\nu$), resulting in a modulation depth of $0.1\text{ dB }\mu\text{m}^{-1}$ and a miniaturized footprint of the modulator. In the left-hand region (with $V_D < -1\text{ V}$), the Fermi level (E_F) is lower than half the photon energy ($-\hbar\nu/2$, blue dashed line) and there are no electrons available for the interband transition. In the right-hand region (with $V_D > 3.8\text{ V}$), all electron states in resonance with incident photons ($\hbar\nu$) are occupied, and the transition is forbidden. In both of the last two cases, the transmission increases. The natural doping from the substrate offsets the centre of the absorption curve from zero bias. The transmission is measured at a laser wavelength of $1.53\text{ }\mu\text{m}$.

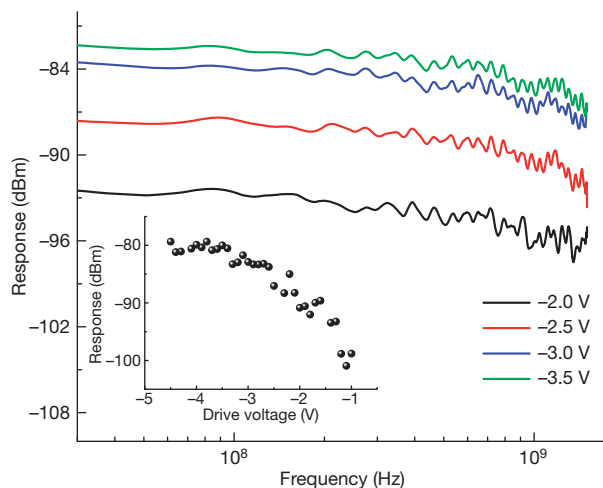


Figure 3 | Dynamic electro-optical response of the device. The main panel shows the response of the device as a function of frequency. Each curve corresponds to a different drive voltage: -2.0 V, -2.5 V, -3.0 V and -3.5 V (black, red, blue and green curve, respectively). The measured 3 dB bandwidths of the device are respectively 0.8 GHz, 1.1 GHz, 1.1 GHz and 1.2 GHz; the bandwidths are mainly restricted by the parasitic response of the device. Inset, low-frequency device response with different drive voltages, indicating that the device has best performance at a drive voltage of -4 V. The laser wavelength is 1.53 μm in the test.

at various drive voltages is obtained. Owing to the exceptionally high carrier mobility and saturation velocity of graphene, the bandwidth is not limited by the carrier transit time, but by the parasitic response of the device. With the platinum electrode placed 500 nm away from the waveguide, the total resistance of the system is reduced to around $600\ \Omega$. This resistance, together with the capacitance (of the order of 0.22 pF), limits the operation bandwidth of the present device to about 1 GHz (see Fig. 3 legend).

The device response at low frequency (300 kHz) is shown in Fig. 3 inset. At low V_D , the modulation response is weak, as the optical transmission is insensitive to V_D . When the drive voltage is increased, the r.f. response increases to a maximum at $V_D = -4$ V. As the drive voltage increases further, the modulation efficiency saturates, as graphene is then transparent within the modulation range of the bias voltage.

As the overall optical opacity of graphene is independent of wavelength and the high frequency dynamic conductivity for Dirac fermions is constant, the graphene electroabsorption modulator is therefore intrinsically broadband, unlike modulators that are based on optical cavities or resonant optical effects (such as QCSE). In order to investigate this broadband effect, we study the static response of the device to a white light source from a supercontinuous laser. The response is shown as a function of wavelength and V_D in Fig. 4a; we refer to this as a two-dimensional (2D) transmission spectrum. A 3 dB modulation, corresponding to a transmission value of 2 (a.u.) in Fig. 4a, is achieved for a broad band of wavelengths, from 1.35 μm to 1.6 μm . Although a higher modulation depth and broader wavelength range are expected at a higher drive voltage, we chose to use low drive voltage not only to avoid spacer oxide breakdown but also because high drive voltages increase power consumption and violate voltage restrictions in CMOS devices.

Two-dimensional transmission spectra also allow us to determine the electronic band dispersion of the graphene. As the graphene electroabsorption modulation is dictated by the optical transition, $h\nu = 2E_F$, the graphene modulator has a different response at different wavelengths. Operation at higher photon energy ($h\nu$) always requires a larger change in the Fermi level (E_F). The trace of critical drive voltage (V) for maximum transmission change rate, shown as a dashed line in Fig. 4a, is defined by $h\nu = 2E_F = 2\hbar v_F \sqrt{\eta\pi|V + V_0|}$, where v_F is the Fermi velocity, V_0 is the voltage offset caused by natural doping, and

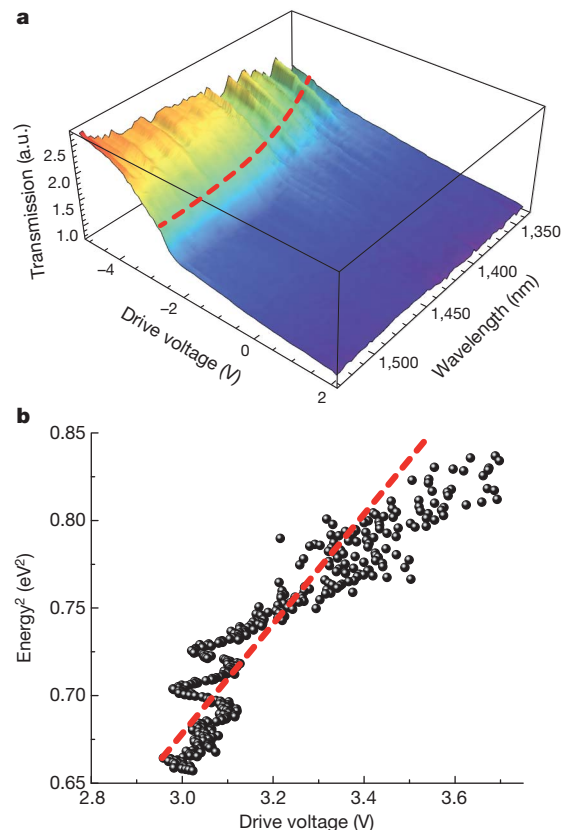


Figure 4 | Spectrum characterization of the optical modulator. **a**, The transmission of the device as a function of drive voltage and wavelength ($1,350$ – $1,600$ nm). The transmission is normalized to $V_D = 1$ V. The red dashed curve shows the trace for maximum transmission change rate, which unambiguously bends to higher drive voltages at shorter wavelengths. **b**, Squared photon energy versus critical drive voltage for maximum transmission change rate. The red dashed line shows a linear fit to the experimental data, which directly yields the Fermi velocity of $v_F = 0.9 \times 10^6$ m s $^{-1}$.

$\eta = 9 \times 10^{16}$ m $^{-2}$ V $^{-1}$, as estimated using a parallel-plate capacitor model of our device. The relation between critical drive voltage and the square of the photon energy is plotted in Fig. 4b. The linear fit (red dashed line) yields the voltage offset (-0.8 V) and the Fermi velocity (0.9×10^6 m s $^{-1}$), which agree with other reported values²⁹.

We have demonstrated a graphene-based optical modulator that has broad optical bandwidth (1.35 – 1.6 μm), small device footprint ($25\ \mu\text{m}^2$) and high operation speed (1.2 GHz at 3 dB) under ambient conditions, all of which are essential for optical interconnects for future integrated optoelectronic systems. The modulation efficiency of a single layer of carbon atoms in a hexagonal lattice (graphene) is already comparable to, if not better than, traditional semiconductor materials such as Si, GeSi and InGaAs, which are orders of magnitude larger in active volume. The flexibility of graphene sheets could also enable radically different photonic devices. For example, graphene could be integrated with flexible substrate and plastic waveguides³⁰. Or it could be used in novel geometries, such as a flexible modulator on a nano-optical cable. The recent development of large scale graphene synthesis and transfer techniques²³ ensure its compatibility with the existing integrated electronics platform.

Received 3 November 2010; accepted 24 March 2011.

Published online 8 May 2011.

1. Miller, D. A. B. Are optical transistors the logical next step? *Nature Photon.* **4**, 3–5 (2010).
2. Reed, G. T., Mashanovich, G., Gardes, F. Y. & Thomson, D. J. Silicon optical modulators. *Nature Photon.* **4**, 518–526 (2010).
3. Liu, A. S. *et al.* A high-speed silicon optical modulator based on a metal-oxide-semiconductor capacitor. *Nature* **427**, 615–618 (2004).

4. Kuo, Y. H. *et al.* Strong quantum-confined Stark effect in germanium quantum-well structures on silicon. *Nature* **437**, 1334–1336 (2005).
5. Liu, J. *et al.* Waveguide-integrated, ultralow-energy GeSi electro-absorption modulators. *Nature Photon.* **2**, 433–437 (2008).
6. Miller, D. A. B. *et al.* Band-edge electroabsorption in quantum well structures — the quantum-confined Stark-effect. *Phys. Rev. Lett.* **53**, 2173–2176 (1984).
7. Xu, Q., Schmidt, B., Pradhan, S. & Lipson, M. Micrometre-scale silicon electro-optic modulator. *Nature* **435**, 325–327 (2005).
8. Novoselov, K. S. Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).
9. Geim, A. K. & Novoselov, K. S. The rise of graphene. *Nature Mater.* **6**, 183–191 (2007).
10. Schwierz, F. Graphene transistors. *Nature Nanotechnol.* **5**, 487–496 (2010).
11. Bonaccorso, F., Sun, Z., Hasan, T. & Ferrari, A. Graphene photonics and optoelectronics. *Nature Photon.* **4**, 611–622 (2010).
12. Liao, L. *et al.* High-speed graphene transistors with a self-aligned nanowire gate. *Nature* **467**, 305–308 (2010).
13. Avouris, P., Chen, Z. H. & Perebeinos, V. Carbon-based electronics. *Nature Nanotechnol.* **2**, 605–615 (2007).
14. Wang, F. *et al.* Gate-variable optical transitions in graphene. *Science* **320**, 206–209 (2008).
15. Li, Z. Q. *et al.* Dirac charge dynamics in graphene by infrared spectroscopy. *Nature Phys.* **4**, 532–535 (2008).
16. Xia, F. N., Mueller, T., Lin, Y. M., Valdes-Garcia, A. & Avouris, P. Ultrafast graphene photodetector. *Nature Nanotechnol.* **4**, 839–843 (2009).
17. Xu, X., Gabor, N. M., Alden, J. S., van der Zande, A. M. & McEuen, P. L. Photo-thermoelectric effect at a graphene interface junction. *Nano Lett.* **10**, 562–566 (2010).
18. Nair, R. R. *et al.* Fine structure constant defines visual transparency of graphene. *Science* **320**, 1308 (2008).
19. Mak, K. F. *et al.* Measurement of the optical conductivity of graphene. *Phys. Rev. Lett.* **101**, 196405 (2008).
20. Bolotin, K. I. *et al.* Ultrahigh electron mobility in suspended graphene. *Solid State Commun.* **146**, 351–355 (2008).
21. Du, X., Skachko, I., Barker, A. & Andrei, E. Y. Approaching ballistic transport in suspended graphene. *Nature Nanotechnol.* **3**, 491–495 (2008).
22. Kampfrath, T., Perfetti, L., Schapper, F., Frischkorn, C. & Wolf, M. Strongly coupled optical phonons in the ultrafast dynamics of the electronic energy and current relaxation in graphite. *Phys. Rev. Lett.* **95**, 187403 (2005).
23. Kim, K. S. *et al.* Large-scale pattern growth of graphene films for stretchable transparent electrodes. *Nature* **457**, 706–710 (2009).
24. Bae, S. *et al.* Roll-to-roll production of 30-inch graphene films for transparent electrodes. *Nature Nanotechnol.* **5**, 574–578 (2010).
25. Reina, A. *et al.* Large area, few-layer graphene films on arbitrary substrates by chemical vapor deposition. *Nano Lett.* **9**, 30–35 (2009).
26. Li, X. *et al.* Large-area synthesis of high-quality and uniform graphene films on copper foils. *Science* **324**, 1312–1314 (2009).
27. Zhang, Y. B., Brar, V. W., Girit, C., Zettl, A. & Crommie, M. F. Origin of spatial charge inhomogeneity in graphene. *Nature Phys.* **5**, 722–726 (2009).
28. Jablan, M., Buljan, H. & Soljačić, M. Plasmonics in graphene at infrared frequencies. *Phys. Rev. B* **80**, 245435 (2009).
29. Zhou, S. Y. *et al.* First direct observation of Dirac fermions in graphite. *Nature Phys.* **2**, 595–599 (2006).
30. Rogers, J. A., Someya, T. & Huang, Y. G. Materials and mechanics for stretchable electronics. *Science* **327**, 1603–1607 (2010).

Acknowledgements This work was supported by the National Science Foundation Nano-scale Science and Engineering Center (NSF-NSEC) for Scalable and Integrated Nano Manufacturing (SINAM) (grant no. CMMI-0751621) and by the US Department of Energy, Basic Energy Sciences Energy Frontier Research Center (DoE-LMI-EFRC) under award DOE DE-AC02-05CH11231. M.L. thanks Y. Rao for discussions.

Author Contributions M.L. and X.Z. contributed to the experimental ideas. M.L. fabricated device samples. M.L. and X.Y. carried out measurements, analysed the experimental data and prepared the manuscript. B.G., L.J. and F.W. prepared graphene film. All authors contributed to discussions and manuscript revision.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to X.Z. (xiang@berkeley.edu, for general experimental details) and F.W. (fengwang76@berkeley.edu, for details of graphene synthesis).

A hydrothermal origin for isotopically anomalous cap dolostone cements from south China

Thomas F. Bristow^{1†}, Magali Bonifacie^{1,2}, Arkadiusz Derkowski³, John M. Eiler¹ & John P. Grotzinger¹

The release of methane into the atmosphere through destabilization of clathrates is a positive feedback mechanism capable of amplifying global warming trends that may have operated several times in the geological past^{1–3}. Such methane release is a hypothesized cause or amplifier for one of the most drastic global warming events in Earth history, the end of the Marinoan ‘snowball Earth’ ice age, ~635 Myr ago^{4–7}. A key piece of evidence supporting this hypothesis is the occurrence of exceptionally depleted carbon isotope signatures ($\delta^{13}\text{C}_{\text{PDB}}$ down to -48‰ ; ref. 8) in post-glacial cap dolostones (that is, dolostone overlying glacial deposits) from south China; these signatures have been interpreted as products of methane oxidation at the time of deposition^{5,6,8}. Here we show, on the basis of carbonate clumped isotope thermometry, $^{87}\text{Sr}/^{86}\text{Sr}$ isotope ratios, trace element content and clay mineral evidence, that carbonates bearing the ^{13}C -depleted signatures crystallized more than 1.6 Myr after deposition of the cap dolostone. Our results indicate that highly ^{13}C -depleted carbonate cements grew from hydrothermal fluids and suggest that their carbon isotope signatures are a consequence of thermogenic methane oxidation at depth. This finding not only negates carbon isotope evidence for methane release during Marinoan deglaciation in south China, but also eliminates the only known occurrence of a Precambrian sedimentary carbonate with highly ^{13}C -depleted signatures related to methane oxidation in a seep environment. We propose that the

capacity to form highly ^{13}C -depleted seep carbonates, through biogenic anaerobic oxidation of methane using sulphate, was limited in the Precambrian period by low sulphate concentrations in sea water⁹. As a consequence, although clathrate destabilization may or may not have had a role in the exit from the ‘snowball’ state, it would not have left extreme carbon isotope signals in cap dolostones.

A common sedimentary motif marking the end of the severe Marinoan ice age (~635 Myr ago¹⁰) is observed in rocks on almost all of the present day continents. Glaciogenic deposits, formed at equatorial latitudes in some places, are sharply overlain by metre-scale intervals of dolostone¹¹. These cap dolostones contain enigmatic sedimentary structures and unusual stable carbon and sulphur isotope signatures hypothesized to reflect climate change associated with deglaciation¹¹. Proposed drivers of global warming during this period include deep-ocean CO_2 outgassing during post-glacial ocean overturn¹², ice–albedo feedback¹¹ or methane release caused by the destabilization of clathrates⁴.

The lithologies and sedimentary structures observed in the cap dolostone of south China, which forms the basal 3–5 m of the Doushantuo Formation, are typical of cap dolostones worldwide^{6,11} (Supplementary Discussion). But unlike other cap dolostones that have mildly negative $\delta^{13}\text{C}_{\text{PDB}}$ values (-2 to -4‰ ; ref. 11), three sections in the Yangtze Gorges area host highly ^{13}C -depleted calcites ($\delta^{13}\text{C}_{\text{PDB}}$ down to -48‰ ; ref. 8). Petrographic textures of the highly ^{13}C -depleted calcites

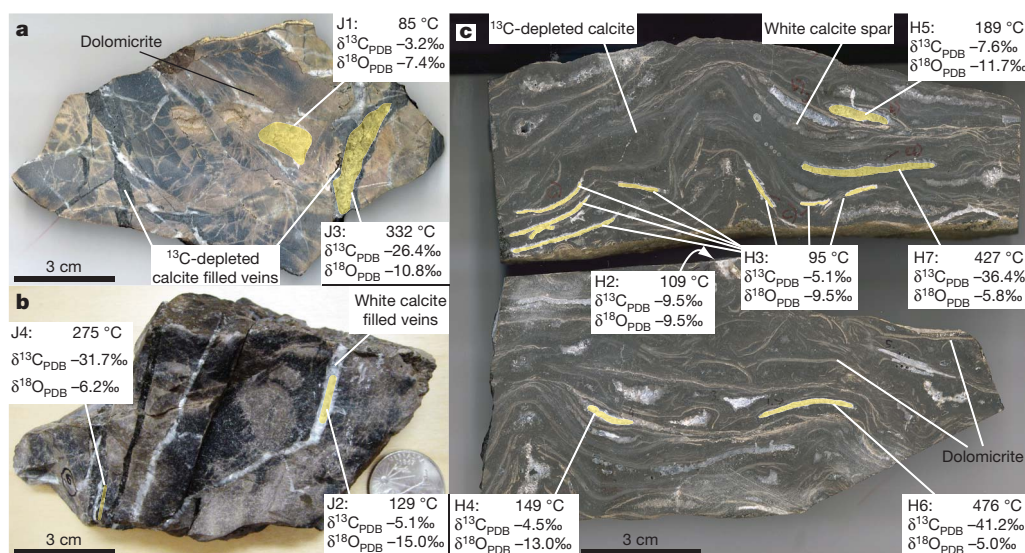


Figure 1 | Crystallization temperatures (based on Δ_{47} measurements) and C and O stable isotope data from various carbonate phases in two samples of the Doushantuo Formation cap dolostone. **a**, **b**, Two views of sample 1, collected from the lower unit of the cap at Jiulongwan (Fig. 3; Supplementary Fig. 1). **c**, Two pieces of sample 3, collected from the middle of the cap dolostone at Huajipo (Fig. 3; Supplementary Fig. 1). The highest temperatures come

exclusively from calcite cements with exceptionally low $\delta^{13}\text{C}_{\text{PDB}}$ values. Areas drilled for isotopic analysis are highlighted in yellow (Supplementary Fig. 2 shows unmarked images of the samples) and labelled (Jx and Hy refer to sample spots from Jiulongwan and Huajipo, respectively) for cross-referencing with data in Supplementary Table 1. The temperatures and isotope values shown are the mean of two or three replicate measurements of the same powder.

¹Division of Geological and Planetary Sciences, California Institute of Technology, 1200 East California Boulevard, Pasadena, California 91125, USA. ²Équipe de Géochimie des Isotopes Stables, Institut de Physique du Globe de Paris, Sorbonne Paris Cité, Université Paris Diderot, UMR 7154 CNRS, F-75005 Paris, France. ³Institute of Geological Sciences, Polish Academy of Sciences, Senacka 1, 31-002 Kraków, Poland. [†]Present address: NASA Ames Research Center, Moffett Field, California 94035, USA.

and associated carbonates have been likened to modern methane seeps^{5,6,8}. Additionally, $\delta^{18}\text{O}_{\text{PDB}}$ values as high as -4‰ in some of the most ^{13}C -depleted calcites are close to values expected in equilibrium with sea water at Earth-surface temperatures, and have been used to support the hypothesis that methane oxidation occurred at the time of deposition⁵.

To test this interpretation, we examined the paragenetic history of three representative samples that contain, amongst other phases, highly ^{13}C -depleted calcite. The samples were collected from the lower and middle units of the cap dolostone from two sections of the Doushantuo Formation in the Yangtze Gorges area at Jiulongwan and Huajipo (see below; Supplementary Fig. 1). Samples consist of: (1) early dolomicrite; (2) highly ^{13}C -depleted grey calcite, with crystals up to 1 mm in length, filling fractures that cross-cut dolomicrite and forming isopachous cements lining voids and bed-parallel lenses; and (3) late-stage white calcite spar filling voids and veins cross-cutting ^{13}C -depleted calcite (Fig. 1; Supplementary Figs 2–6).

Application of carbonate clumped isotope thermometry to these materials provides new constraints on their origin and diagenetic history. This technique measures the degree of ‘clumping’ of heavy isotopes of carbon and oxygen (^{13}C and ^{18}O , respectively) in the carbonate lattice in comparison with a random distribution. The degree of clumping, expressed as Δ_{47} in units of ‰ , shows a systematic dependence on temperature¹³. This thermometer can be used to reconstruct the temperature of carbonate precipitation and the oxygen isotope composition of the fluids ($\delta^{18}\text{O}_{\text{SMOW}}$) from which analysed carbonates formed.

The cap samples examined in this study have Δ_{47} values ranging from 0.487‰ to 0.265‰ , corresponding to temperatures of $86\text{--}476\text{ °C}$ (Figs 1 and 2a; Supplementary Table 1). In all samples, dolomicrite records the lowest temperatures (mean, 112 °C ; $n = 4$), whereas the strongly ^{13}C -depleted grey calcite has the highest temperatures (mean, 378 °C ; $n = 4$), with white calcite spar showing intermediate temperatures (mean, 156 °C ; $n = 3$). The possibility that high carbonate clumped isotope temperatures result from nonlinear mixing or kinetic effects has been considered and ruled out (Supplementary Discussion). Further discussion of data from dolomicrite and white calcite spar is provided in the Supplementary Discussion; here we focus on the origin of highly ^{13}C -depleted calcite.

Carbonate clumped isotope temperatures from highly ^{13}C -depleted calcite are up to $200\text{--}300\text{ °C}$ higher than in any natural carbonate sample analysed to date. Even marbles that experienced $\sim 500\text{ °C}$ during regional metamorphism have lower apparent carbonate clumped temperatures ($\sim 200\text{ °C}$) because cooling over millions of years allowed C–O bonds to keep rearranging until a final ‘blocking’ temperature was reached¹³. Preservation of high temperatures in the Doushantuo Formation therefore requires faster cooling rates, which we propose are related to a local, short-lived thermal anomaly. More specifically, we suggest that highly ^{13}C -depleted calcite precipitated from a pulse of hot hydrothermal fluid. The exact dependence of blocking temperature on cooling rate is currently unknown. The only potentially relevant constraint is that rearrangement of oxygen by self-diffusion at 500 °C over length scales similar to a unit cell is likely to occur in minutes¹⁴. Conversely, carbonate blocking temperatures of $\sim 200\text{ °C}$ show that the rate of oxygen diffusion drops rapidly with decreasing temperature. Therefore, the lowest temperatures measured in highly ^{13}C -depleted calcite ($\sim 275\text{ °C}$) could have persisted over geological timescales.

A series of independent observations are consistent with a post-depositional, hydrothermal origin for the highly ^{13}C -depleted calcite. First, by combining temperatures with $\delta^{18}\text{O}_{\text{PDB}}$ values of carbonate, we calculate that ^{13}C -depleted calcite precipitated from ^{18}O -enriched fluids ($\delta^{18}\text{O}_{\text{SMOW}} \geq +18\text{‰}$), which were distinct from the fluids that precipitated dolomicrite and calcite spar ($\delta^{18}\text{O}_{\text{SMOW}}$ of $1 \pm 1\text{‰}$ and $13 \pm 2\text{‰}$, respectively; Fig. 2b, Supplementary Table 1). ^{18}O -enrichment of fluids in seep environments can be caused by clathrate dissociation^{7,15}. However, the maximum oxygen isotope fractionation during clathrate formation is $+3.2\text{‰}$ relative to the water source¹⁶; therefore clathrate

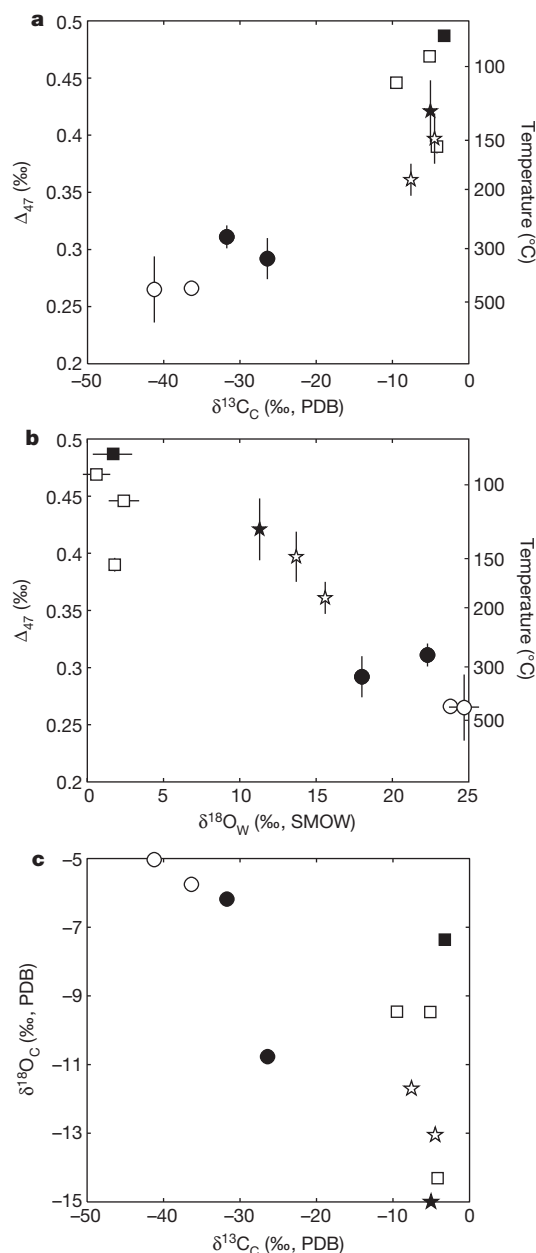


Figure 2 | Cross-plots of Δ_{47} and conventional stable isotope data from the three carbonate phases identified in cap dolostone samples. **a**, Mean Δ_{47} (left axis) and derived temperature (right axis) versus $\delta^{13}\text{C}_{\text{PDB}}$ of carbonate ($\delta^{13}\text{C}_{\text{C}}$). **b**, Mean Δ_{47} (left axis) and derived temperature (right axis) versus the calculated mean $\delta^{18}\text{O}_{\text{SMOW}}$ of water precipitating carbonates ($\delta^{18}\text{O}_{\text{W}}$). **c**, Mean $\delta^{18}\text{O}_{\text{PDB}}$ of carbonate ($\delta^{18}\text{O}_{\text{C}}$) versus $\delta^{13}\text{C}_{\text{PDB}}$ of carbonate. Error bars, ± 1 s.d. based on two or more replicate analyses of the same powder. Error bars are less than the width and height of the symbol where not apparent. Open symbols, samples from the Huajipo section; filled symbols, samples from Jiulongwan. Stars, white calcite spar; squares, dolomicrite; and circles, ^{13}C -depleted grey calcite.

dissociation cannot account for the degree of $\delta^{18}\text{O}_{\text{SMOW}}$ enrichment calculated for fluids precipitating ^{13}C -depleted calcites. Conversely, high- $\delta^{18}\text{O}_{\text{SMOW}}$ fluids (similar to those that precipitated ^{13}C -depleted calcite) are reported from other ancient continental hydrothermal systems where ^{18}O -rich carbonate host rocks control oxygen isotope fluid compositions¹⁷. Therefore, we think that the high $\delta^{18}\text{O}_{\text{PDB}}$ values observed in ^{13}C -depleted calcite (Fig. 2c) are not a sign of exceptional preservation (as previously suggested, ref. 5), but instead are a result of precipitation from hot fluids buffered by ^{18}O -enriched carbonate host rocks (i.e. at low water/rock ratios).

Second, elemental analysis shows that highly ^{13}C -depleted calcite has Mn/Sr ratios >100 (Supplementary Table 2). These cements also have $^{87}\text{Sr}/^{86}\text{Sr}$ values (0.7090 to 0.7130) that are radiogenic in comparison with the best-preserved Marinoan cap dolostones (0.7072 to 0.7080)¹⁸ and the low Mn/Sr (<1) carbonates from argillaceous dolomites overlying the Doushantuo Formation cap (~ 0.7080 , ref. 19; Fig. 3, Supplementary Table 3). In contrast, most well preserved Phanerozoic seep carbonates have $^{87}\text{Sr}/^{86}\text{Sr}$ values close to contemporaneous sea water and Mn/Sr ratios <1 (Supplementary Tables 4 and 5). However, it is noteworthy that dolomicrite and white calcite spar also have relatively high Mn/Sr and $^{87}\text{Sr}/^{86}\text{Sr}$ ratios (Supplementary Table 2), indicating they were diagenetically altered²⁰. This raises the possibility that the Mn/Sr and $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of ^{13}C -depleted calcite may not be uniquely diagnostic of its origin (Supplementary Discussion). And it also makes conceivable the preservation of depositional $\delta^{13}\text{C}_{\text{PDB}}$ values in highly ^{13}C -depleted calcite, if Δ_{47} were to be reset during exchange of isotopes and trace elements between pre-existing carbonate and hot, carbon-poor fluid with high $\delta^{18}\text{O}$, high Mn and high $^{87}\text{Sr}/^{86}\text{Sr}$. But such a process strikes us as unlikely, because it should have affected other phases in the cap dolostone rather than being exclusive to highly ^{13}C -depleted calcite.

Highly ^{13}C -depleted diagenetic carbonates (with $\delta^{13}\text{C}_{\text{PDB}}$ values down to -41‰), interpreted as a product of thermogenic oxidation of low- $\delta^{13}\text{C}_{\text{PDB}}$ hydrocarbons such as methane, have been reported from several other sedimentary basins^{21,22}. Similarly, we infer that the extremely low- $\delta^{13}\text{C}_{\text{PDB}}$ calcite in Doushantuo formed via thermochemical oxidation of hydrothermal methane. We think methane was sourced from organic-rich marls of the Doushantuo Formation, as there are no older source rocks in the Yangtze Gorges area. This is an appealing hypothesis, because the most ^{13}C -depleted methane (down to -51‰ , PDB) previously observed in continental hydrothermal systems is produced by the thermogenic breakdown of organic matter in host sediments²³. In addition, we suggest that the low permeability of overlying clay-rich lithologies of the Doushantuo Formation acted as a seal, causing preferential hydrothermal fluid flow through the porous cap dolostone. This explains why highly ^{13}C -depleted carbonates have only been reported from the cap.

Systematic variations in the degree of clay mineral diagenesis indicate a localized thermal anomaly at the base of the Doushantuo Formation, thus supporting the hypothesis of focused hydrothermal flow. The main clay mineral in the lower 80 m of the Doushantuo Formation in this area is saponite, an Mg-rich trioctahedral smectite interpreted as forming at the time of deposition, based on the unusual mineralogy²⁴. However, X-ray diffraction (XRD) data²⁴ and a decrease in cation exchange capacity (CEC) normalized to the total clay content (Fig. 3; Supplementary Discussion) show that saponite is increasingly altered to corrensite (ordered mixed-layer trioctahedral smectite/chlorite) and chlorite down-section, as the cap dolostone is approached. Because the extent of chloritization of saponite during diagenesis increases with temperature and duration of thermal activity, and does not require extensive fluid exchange, it is a useful means of monitoring maximum palaeotemperatures and is therefore commonly used to study the diagenetic history of sedimentary basins²⁵. At temperatures $>300^\circ\text{C}$, indicated by carbonate clumped isotopes, chloritization takes place at timescales of hundreds to thousands of years²⁵, consistent with our hypothesis that low Δ_{47} values in carbonates were preserved by rapid cooling.

The thermal gradient implied by clay minerals in the basal 25 m of the Jiulongwan section (Fig. 3) also requires that hydrothermal activity took place after deposition of at least this much sediment. U–Pb zircon ages of 635.2 ± 0.5 Myr and 632.5 ± 0.6 Myr from ash beds within the cap dolostone and 5 m above the top of the cap, respectively¹⁰, show that hydrothermal activity occurred more than 1.6 Myr after deposition of the cap dolostone and is therefore unrelated to deglaciation. We suspect a much younger age, however, coinciding with a regionally extensive Early Cambrian hydrothermal episode in south China (ref. 26; Supplementary Discussion).

Our findings show that the highly ^{13}C -depleted calcite in the Doushantuo Formation is not a record of clathrate destabilization associated with Marinoan deglaciation. The results also highlight a broader puzzle. Before this study, the Doushantuo Formation was considered to contain the oldest examples of carbonates derived by methane oxidation at a cold seep, despite evidence for metabolism of methane dating back to the Archaean²⁷ and predictions of higher

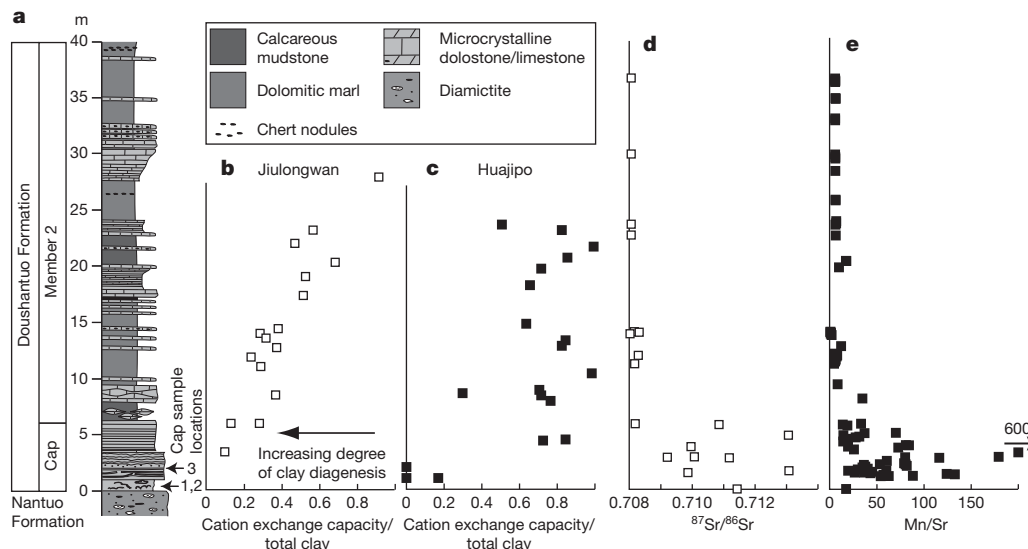


Figure 3 | Stratigraphic variability in trace element content, $^{87}\text{Sr}/^{86}\text{Sr}$ ratios and degree of chloritization of trioctahedral clay minerals of the lower Doushantuo Formation from the two sections examined in this study. **a**, Stratigraphy. **b**, **c**, Stratigraphic trends in the cation exchange capacity (CEC) of a sample normalized to the total trioctahedral clay content. This ratio is used as an indicator of the extent of clay mineral diagenesis (Supplementary Discussion). Samples with lower normalized CEC contain more corrensite and chlorite and are more diagenetically altered. The stratigraphic position of

mineral transformations is different in each section; this is an expected consequence of the heterogeneity of a thermal anomaly induced by hypothesized hydrothermal activity. $^{87}\text{Sr}/^{86}\text{Sr}$ ratios (**d**) and Mn/Sr data (**e**) are from ref. 19 and are consistent with phase specific trace element and $^{87}\text{Sr}/^{86}\text{Sr}$ data from the cap dolostone (Supplementary Tables 2 and 3). The sampling locations of cap dolostone samples analysed are shown next to the stratigraphic column. Sample 1 comes from Jiulongwan, samples 2 and 3 come from Huajipo.

fluxes of biogenic methane in the Precambrian²⁸. However, our interpretation of the highly ¹³C-depleted calcite, combined with a survey of seep occurrences through Earth history¹⁵ and a recent survey of Precambrian carbon isotope data (including over 11,000 analyses)²⁹, show that Precambrian carbonates are devoid of cements with $\delta^{13}\text{C}_{\text{PDB}}$ values less than -30‰ , which are characteristic of methane seep carbonates in the Phanerozoic¹⁵.

We propose that this absence reflects the importance of anaerobic oxidation of methane (AOM), using sulphate, in generating exceptionally low $\delta^{13}\text{C}_{\text{PDB}}$ signatures in carbonate rocks associated with methane seeps. AOM is a biologically mediated reaction that generates ¹³C-depleted carbonate alkalinity, promoting subsequent precipitation of carbonate with exceptionally low $\delta^{13}\text{C}_{\text{PDB}}$. Incubation experiments on cold seep sediments, naturally enriched in methanotrophic communities, show that decreased sulphate concentrations result in reduced rates of AOM³⁰. Therefore, low sulphate concentrations that characterized Precambrian oceans⁹ would have reduced AOM rates, making conditions less favourable for the precipitation of highly ¹³C-depleted carbonates¹¹. Our reinterpretation of the highly ¹³C-depleted carbonates in the Doushantuo Formation thus highlights the influence of ocean chemistry on methane cycling through Earth history.

METHODS SUMMARY

We characterized the petrography and mineralogy of three cap dolostone samples using visible light microscopy, elemental mapping (with the electron microprobe JEOL JXA 8200 at the California Institute of Technology and an XGT X-ray fluorescence scanner at the Jet Propulsion Laboratory) and quantitative elemental spot analysis of carbonates by electron microprobe. Representative carbonates were micro-drilled from slabs for isotope measurements. Sr isotope measurements were made on acetic acid digestions of carbonate powders using a Neptune MC-ICPMS at the Keck Laboratory, University California, Santa Cruz. Clumped and traditional carbon and oxygen isotope analysis were made on a MAT 253 at the California Institute of Technology. The degree of chloritization of clay minerals was quantified by measuring the CEC to total clay content ratios of 35 samples of the cap and 25 m of overlying sediments collected at Huajipo and Jiulongwan. Quantitative XRD and CEC measurements were carried out at the University of California, Riverside, on whole-rock powders (see Methods and Supplementary Discussion for more details).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 August 2010; accepted 25 March 2011.

Published online 25 May 2011.

- MacDonald, G. J. Role of methane clathrates in past and future climates. *Clim. Change* **16**, 247–281 (1990).
- Dickens, G. R., O'Neil, J. R., Rea, D. K. & Owen, R. M. Dissociation of oceanic methane hydrate as a cause of the carbon isotope excursion at the end of the Paleocene. *Paleoceanography* **10**, 965–971 (1995).
- Hesselbo, S. P. *et al.* Massive dissociation of gas hydrate during a Jurassic oceanic anoxic event. *Nature* **406**, 392–395 (2000).
- Kennedy, M. J., Christie-Blick, N. & Sohl, L. E. Are Proterozoic cap carbonates and isotopic excursions a record of gas hydrate destabilization following Earth's coldest intervals? *Geology* **29**, 443–446 (2001).
- Jiang, G. Q., Kennedy, M. J. & Christie-Blick, N. Stable isotopic evidence for methane seeps in Neoproterozoic postglacial cap carbonates. *Nature* **426**, 822–826 (2003).
- Jiang, G. Q., Kennedy, M. J., Christie-Blick, N., Wu, H. C. & Zhang, S. H. Stratigraphy, sedimentary structures, and textures of the late Neoproterozoic Doushantuo cap carbonate in south China. *J. Sedim. Res.* **76**, 978–995 (2006).
- Kennedy, M., Mrofka, D. & von der Borch, C. Snowball Earth termination by destabilization of equatorial permafrost methane clathrate. *Nature* **453**, 642–645 (2008).
- Wang, J. S., Jiang, G. Q., Xiao, S. H., Li, Q. & Wei, Q. Carbon isotope evidence for widespread methane seeps in the ca. 635 Ma Doushantuo cap carbonate in south China. *Geology* **36**, 347–350 (2008).

- Kah, L. C., Lyons, T. W. & Frank, T. D. Low marine sulphate and protracted oxygenation of the Proterozoic biosphere. *Nature* **431**, 834–838 (2004).
- Condon, D. *et al.* U–Pb ages from the Neoproterozoic Doushantuo Formation, China. *Science* **308**, 95–98 (2005).
- Hoffman, P. F. & Schrag, D. P. The snowball Earth hypothesis: testing the limits of global change. *Terra Nova* **14**, 129–155 (2002).
- Grotzinger, J. P. & Knoll, A. H. Anomalous carbonate precipitates; is the Precambrian the key to the Permian? *Palaios* **10**, 578–596 (1995).
- Ghosh, P. *et al.* ¹³C–¹⁸O bonds in carbonate minerals: a new kind of paleothermometer. *Geochim. Cosmochim. Acta* **70**, 1439–1456 (2006).
- Farver, J. R. Oxygen self-diffusion in calcite: dependence on temperature and water fugacity. *Earth Planet. Sci. Lett.* **121**, 575–587 (1994).
- Campbell, K. A. Hydrocarbon seep and hydrothermal vent paleoenvironments and paleontology: past developments and future research directions. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **232**, 362–407 (2006).
- Maekawa, T. Experimental study on isotopic fractionation in water during gas hydrate formation. *Geochem. J.* **38**, 129–138 (2004).
- Bechtel, A., Savin, S. M. & Hoernes, S. Oxygen and hydrogen isotopic composition of clay minerals of the Bahloul Formation in the region of the Bou Grine zinc-lead ore deposit (Tunisia): evidence for fluid–rock interaction in the vicinity of salt dome cap rock. *Chem. Geol.* **156**, 191–207 (1999).
- Halverson, G. P., Dudas, F. O., Maloof, A. C. & Bowring, S. A. Evolution of the ⁸⁷Sr/⁸⁶Sr composition of Neoproterozoic seawater. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **256**, 103–129 (2007).
- Sawaki, Y. *et al.* The Ediacaran radiogenic Sr isotope excursion in the Doushantuo Formation in the Three Gorges area, South China. *Precamb. Res.* **176**, 46–64 (2010).
- Jacobsen, S. B. & Kaufman, A. J. The Sr, C and O isotopic evolution of Neoproterozoic seawater. *Chem. Geol.* **161**, 37–57 (1999).
- Boles, J. R., Eichhubl, P., Garven, G. & Chen, J. Evolution of a hydrocarbon migration pathway along basin-bounding faults: evidence from fault cement. *Bull. Am. Assoc. Petrol. Geol.* **88**, 947–970 (2004).
- Machel, H. G., Cavell, P. A. & Patey, K. S. Isotopic evidence for carbonate cementation and recrystallization, and for tectonic expulsion of fluids into the Western Canada Sedimentary Basin. *Geol. Soc. Am. Bull.* **108**, 1108–1119 (1996).
- Welhan, J. A. Origins of methane in hydrothermal systems. *Chem. Geol.* **71**, 183–198 (1988).
- Bristow, T. F. *et al.* Mineralogical constraints on the paleoenvironments of the Ediacaran Doushantuo Formation. *Proc. Natl Acad. Sci. USA* **106**, 13190–13195 (2009).
- Meunier, A. *Clays* (Springer, 2005).
- Chen, D., Wang, J., Qing, H., Yan, D. & Li, R. Hydrothermal venting activities in the Early Cambrian, South China: petrological, geochronological and stable isotope constraints. *Chem. Geol.* **258**, 168–181 (2009).
- Eigenbrode, J. L. & Freeman, K. H. Late Archean rise of aerobic microbial ecosystems. *Proc. Natl Acad. Sci. USA* **103**, 15759–15764 (2006).
- Kasting, J. F. Methane and climate during the Precambrian era. *Precamb. Res.* **137**, 119–129 (2005).
- Knauth, L. P. & Kennedy, M. J. The late Precambrian greening of the Earth. *Nature* **460**, 728–732 (2009).
- Wegener, G. & Boetius, A. An experimental study on short-term changes in the anaerobic oxidation of methane in response to varying methane and sulfate fluxes. *Biogeosciences* **6**, 867–876 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Jiang for guidance in the field and for providing samples, V. Orphan and W. Fischer for discussion and advice, M. Kennedy for supporting fieldwork and use of analytical equipment, and E. Peterman and K. Morrison for laboratory work. C. Ma and M. Anderson are thanked for analytical assistance and advice. This work was supported by an O.K. Earl Postdoctoral fellowship (to T.F.B.), by the NSF EAR and GEG programmes (to J.M.E.), and by INSU (to M.B.). Part of the work of M.B. is IGP contribution 3138.

Author Contributions T.F.B. and J.P.G. conceived the study. M.B. carried out clumped and conventional isotope analysis and wrote part of the Supplementary Discussion. J.M.E. provided laboratory facilities for isotope work. A.D. and T.F.B. carried out clay mineral analysis. T.F.B. carried out petrographic work and trace element analysis and wrote the manuscript. All authors discussed results, planned analyses and contributed to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.F.B. (thomas.f.bristow@nasa.gov).

METHODS

Clumped and traditional isotope measurements. CO₂ was extracted from all carbonate samples by phosphoric acid digestion at 90 °C using the automatic laboratory methods described elsewhere³¹. CO₂ was analysed at the California Institute of Technology using a Finnigan MAT 253 gas source mass spectrometer configured to collect masses 44–49. Each measurement consisted of eight acquisitions, with typical standard deviations of 0.01‰ to 0.04‰ for Δ_{47} measurements. Values of $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ were acquired as part of each analysis and typically show standard deviations one order of magnitude lower (averages of 0.008‰ and 0.004‰, respectively). Carbonate powders with known compositions and heated CO₂ standards were run with unknown samples for offset correction and standardization. Heated gas (CO₂ heated for two hours at 1,000 °C to achieve a stochastic isotopic distribution) with a range of bulk stable isotope compositions similar to samples ($\delta^{13}\text{C} \ll 0\text{‰}$) were analysed to minimize the potential errors associated with mass spectrometric nonlinearities. For this study, data were normalized relative to a fixed heated gas line model (for each of the three sessions of analyses run) and raw data corrected for instrument nonlinearity and scale compression as described in ref. 32. Δ_{47} data were then corrected for an acid reaction temperature of 90 °C, with the correction factor of 0.081‰, as experimentally determined³¹. Finally, Δ_{47} data were normalized to working carbonate standards: a vein calcite named 102-GC-AZ01 and an Italian marble named Carrara marble. This last correction stage is typically less than 0.02‰. Fourteen distinct extractions of Carrara marble standard during this period yielded a mean Δ_{47} value of $0.356 \pm 0.012\text{‰}$, a $\delta^{13}\text{C}_{\text{PDB}}$ of $2.34 \pm 0.04\text{‰}$ and a $\delta^{18}\text{O}_{\text{PDB}}$ of $-1.75 \pm 0.07\text{‰}$. Five separate extractions of 102-GC-AZ01 yielded a Δ_{47} of $0.646 \pm 0.009\text{‰}$, a $\delta^{13}\text{C}_{\text{PDB}}$ of $0.51 \pm 0.06\text{‰}$ and a $\delta^{18}\text{O}_{\text{PDB}}$ of $-14.32 \pm 0.06\text{‰}$. Accepted Δ_{47} values for these standards from >60 analyses by multiple analysts in our laboratory are 0.352‰ and 0.654‰ for Carrara marble and 102-GC-AZ01, respectively.

Fractionation factors of 1.00821 and 1.0093 were used to account for the temperature-dependent oxygen isotope fractionation between CO₂ gas and carbonates resulting from the reaction with phosphoric acid at 90 °C, for calcitic samples³³ and dolomitic samples³⁴, respectively. Measured values of Δ_{47} (in ‰) were used to estimate carbonate growth temperature, using the empirically derived polynomial determined using high temperature experimental carbonates¹³, hydrothermal dolomite (M.B. *et al.*, manuscript in preparation) and inorganic synthetic calcites¹³. Paired temperature and carbonate $\delta^{18}\text{O}_{\text{PDB}}$ data were used to calculate the $\delta^{18}\text{O}_{\text{SMOW}}$ values of hydrothermal and/or formation waters that interacted with analysed carbonates, using temperature dependent carbonate–water fractionations described in ref. 35 for calcite and ref. 36 for dolomite. All samples were analysed at least twice using sub-fractions of the same powder, to account for heterogeneity. Uncertainties in temperature estimates and isotopic data (Fig. 2; Supplementary Table 1) are based on a minimum of two analyses of the same powder. A minimum error for Δ_{47} measurements of 0.02‰ has been applied based on the maximum external precision expected from counting statistics; see, for example, ref. 32.

Sr ratios. Because of the relatively large amount of sample required for combined Δ_{47} and Sr isotope analysis, powders for Sr isotope analysis of sample 3 were collected from the same phases in a companion slab cut from the sample used

to obtain powders for isotope measurements. For other samples, powders were obtained by re-drilling holes initially used to collect powder for Δ_{47} measurements. To avoid potential contamination from silicate phases, Sr isotope measurements were made on acetic acid digests of carbonate powders, using a Neptune MC-ICPMS at the Keck Laboratory, University California, Santa Cruz. The precision of these measurements is 0.00002, based on repeated measurements of an internal standard that yielded a mean Sr ratio of 0.71030 ($n = 5$). The accepted value of this standard is 0.71025 and this offset was applied in correcting measured sample ratios. A repeat measurement of a sample of highly ¹³C-depleted calcite from sample 3 yielded Sr ratios that were within 0.00002.

Elemental mapping and analysis. Two methods were used in elemental mapping. Areas of samples 1 and 3, outlined in blue in Supplementary Fig. 4 and 5, were scanned using an XGT X-ray fluorescence scanner at the Jet Propulsion Laboratory, Pasadena, California. The images give qualitative information about the elemental abundances in various phases of the samples, with the brightest areas containing the highest relative abundance of a particular element.

A highly polished, large format (51 × 75 mm) thin section, of part of sample 3, corresponding to the area inside the blue box shown in Supplementary Fig. 5, was carbon coated for electron probe analysis using the JEOL JXA 8200 instrument at the Division of Geological and Planetary Sciences Analytical Facility, California Institute of Technology. A representative 6 × 6 mm area that contained each of the three main carbonate phases was selected for elemental mapping. The elemental maps shown in Supplementary Fig. 6 give qualitative information about elemental abundance, with brighter areas containing higher concentrations of an element.

Quantitative measurements of Mg, Ca, Sr, Fe, Mn, Si, Al and Ba in the various carbonate phases were made using multiple spot analyses, with a defocused 10 µm spot at 15 kV and 15 nA.

Quantitative XRD and CEC measurements. These analyses were made at the University of California, Riverside, on powdered samples from the same sections of the Doushantuo Formation (at Huajipo and Jiulongwan) that cap dolostone samples were collected from (Fig. 3). The methods used for analysis are described elsewhere^{24,37}.

- Passey, B. H., Levin, N. E., Cerling, T. E., Brown, F. H. & Eiler, J. M. High-temperature environments of human evolution in East Africa based on bond ordering in paleosol carbonates. *Proc. Natl Acad. Sci. USA* **107**, 11245–11249 (2010).
- Huntington, K. W. *et al.* Methods and limitations of 'clumped' CO₂ isotope (Δ_{47}) analysis by gas-source isotope ratio mass spectrometry. *J. Mass Spectrom.* **44**, 1318–1329 (2009).
- Swart, P. K., Burns, S. J. & Leder, J. J. Fractionation of the stable isotopes of oxygen and carbon in carbon-dioxide during the reaction of calcite with phosphoric acid as a function of temperature and technique. *Chem. Geol.* **86**, 89–96 (1991).
- Rosenbaum, J. & Sheppard, S. M. F. An isotopic study of siderites, dolomites and ankerites at high-temperatures. *Geochim. Cosmochim. Acta* **50**, 1147–1150 (1986).
- Friedman, I. & O'Neil J. R. Compilation of stable isotope fractionation factors of geochemical interest. *Prof. Pap. US Geol. Surv.* **440-KK**, (1977).
- Horita, J. The dolomite problem: oxygen isotope fractionation to elevated temperatures. *Geochim. Cosmochim. Acta* **72**, A391 (2008).
- Orsini, L. & Remy, J. C. Utilisation du chlorure de cobalthexamine pour la détermination simultanée de la capacité d'échange et des bases échangeables des sols. *Sci. Sol* **4**, 269–275 (1976).

A dynamic early East Antarctic Ice Sheet suggested by ice-covered fjord landscapes

Duncan A. Young¹, Andrew P. Wright², Jason L. Roberts^{3,4}, Roland C. Warner^{3,4}, Neal W. Young^{3,4}, Jamin S. Greenbaum¹, Dustin M. Schroeder¹, John W. Holt¹, David E. Sugden², Donald D. Blankenship¹, Tas D. van Ommen^{3,4} & Martin J. Siegert²

The first Cenozoic ice sheets initiated in Antarctica from the Gamburtsev Subglacial Mountains¹ and other highlands as a result of rapid global cooling ~34 million years ago². In the subsequent 20 million years, at a time of declining atmospheric carbon dioxide concentrations² and an evolving Antarctic circumpolar current², sedimentary sequence interpretation³ and numerical modelling⁴ suggest that cyclical periods of ice-sheet expansion to the continental margin, followed by retreat to the subglacial highlands, occurred up to thirty times. These fluctuations were paced by orbital changes and were a major influence on global sea levels⁵. Ice-sheet models show that the nature of such oscillations is critically dependent on the pattern and extent of Antarctic topographic lowlands. Here we show that the basal topography of the Aurora Subglacial Basin of East Antarctica, at present overlain by 2–4.5 km of ice, is characterized by a series of well-defined topographic channels within a mountain block landscape. The identification of this fjord landscape, based on new data from ice-penetrating radar, provides an improved understanding of the topography of the Aurora Subglacial Basin and its surroundings, and reveals a complex surface sculpted by a succession of ice-sheet configurations substantially different from today's. At different stages during its fluctuations, the edge of the East Antarctic Ice Sheet lay pinned along the margins of the Aurora Subglacial Basin, the upland boundaries of which are currently above sea level and the deepest parts of which are more than 1 km below sea level. Although the timing of the channel incision remains uncertain, our results suggest that the fjord landscape was carved by at least two ice-flow regimes of different scales and directions, each of which would have over-deepened existing topographic depressions, reversing valley floor slopes.

Deep-sea oxygen isotope records show the onset of significant glaciation in Antarctica at the Eocene/Oligocene boundary^{2,5} (~34 million years (Myr) ago). Morphological evidence for sustained alpine-style glaciation in the Gamburtsev Subglacial Mountains, underlying the Dome A region of the East Antarctic Ice Sheet (EAIS), shows that they were a centre of ice-sheet initiation¹. Although it is thought that the EAIS has remained in a persistent state for the last 14 Myr (as evidenced in the Antarctic Dry Valleys by very low erosion rates⁶, cold-based local glaciers⁷ and the preservation of buried Miocene ice⁸), offshore sedimentary records³ point to there being major oscillations in ice-sheet surface area between 34 and 14 Myr ago. Exactly how these oscillations were expressed by the ice sheet is, however, poorly constrained.

Numerical ice-sheet models can be used to understand the form and flow of past ice sheets. Such models indicate that ice growth begins at higher elevations (such as the Gamburtsev Subglacial Mountains) before encroaching on lower regions^{4,9–11}. The bed elevation grids used as input to these models are, in some regions, constructed from sparse data^{12,13}. One such region is the Aurora Subglacial Basin (ASB; Fig. 1), which from reconnaissance data is known to be a deep trough (more than 1 km below sea level) oriented nearly orthogonal to the modern

ice margin and located to the northeast of the elevated Dome A and Ridge B regions of the ice sheet (Fig. 2a). Ice-sheet models^{10,11} demonstrate the potential importance of the ASB to the progression of ice-sheet growth. These models show a large, growing ice mass from Dome A and Ridge B that converges with smaller radial ice cover from Dome C, resulting in the ASB being buried with deep glacial ice, as at present (Fig. 2c). These models also show that ice-sheet decay is likely to begin in the lowlands of the ASB, isolating a radial ice cap at Dome C and pushing the ice margin back towards Ridge B, eventually depleting the ASB of ice altogether (Fig. 2c). Although it is clear that the ASB has a potentially significant influence on EAIS stability, paucity of bed data, especially around the transition between the ice margin and the interior, is a source of exceptional uncertainty in estimates of the rates and magnitudes of past and present global sea-level changes.

To address this knowledge gap, the ICECAP aerogeophysical programme (Methods) acquired 47,492 line kilometres of airborne radar profiles over the ASB, and from these data a new bed topography has been established (Fig. 2a). The new data extend over a semicircular region radiating from Law Dome and cover approximately $1.5 \times 10^6 \text{ km}^2$ (11% of the Antarctic ice sheet). The region extends from Denman Glacier in

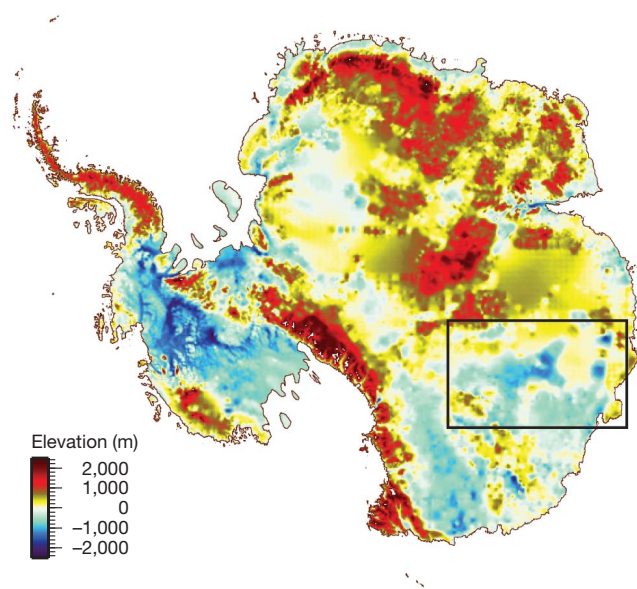


Figure 1 | Bed topography of Antarctica. The blue areas represent Antarctica's major marine subglacial basins. This data set^{12,13} is an interpolation of existing data, which are sparse in the region of the ASB (black box; see also Fig. 2).

¹Institute for Geophysics, Jackson School of Geosciences, University of Texas at Austin, Austin, Texas 78758, USA. ²School of GeoSciences, University of Edinburgh, King's Buildings, Edinburgh EH9 3JW, UK. ³Australian Antarctic Division, Kingston, Tasmania 7050, Australia. ⁴Antarctic Climate & Ecosystems Cooperative Research Centre, Hobart, Tasmania 7001, Australia.

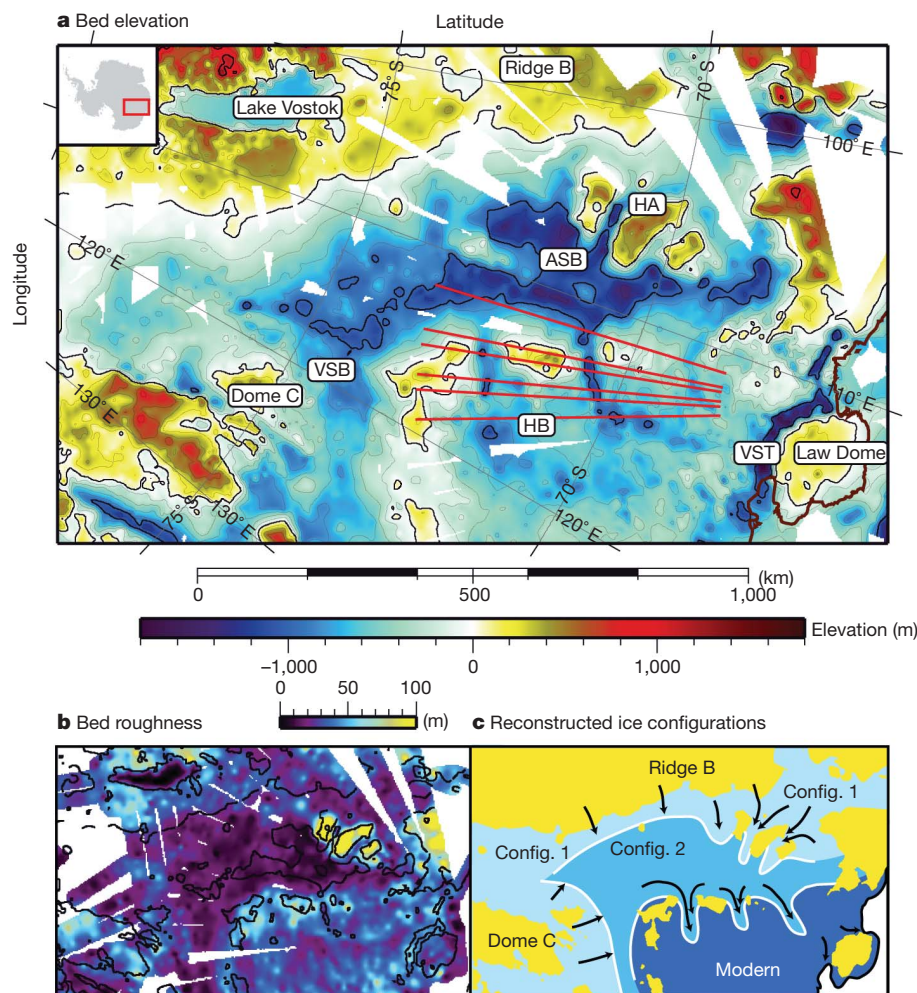


Figure 2 | Bed topography of the ASB region, East Antarctica. **a**, Detailed map derived from radio-echo sounding data from ICECAP, the Support Office for Aerogeophysical Research and BEDMAP, using a natural-neighbour interpolation scheme. Features include subglacial Lake Vostok, Vincennes Subglacial Basin (VSB), Vanderford Subglacial Trench (VST), Law Dome and the ASB. The newly defined features Highland A (HA) and Highland B (HB) are also indicated. See Supplementary Fig. 1 for flight tracks and Supplementary Fig. 2 for data used; regions more than 50 km from data are masked in white.

the west to Dome C in the south and to Porpoise Bay in the east. The deepest point ($-2,426 \pm 10$ m in the WGS-84 coordinate frame) is near the coast in the Vanderford Subglacial Trench, through which both the Vanderford and the Totten Glaciers drain; the highest point ($1,637 \pm 10$ m WGS-84) lies in a previously unknown subglacial mountain range (Highland A) 400 km southeast of Denman Glacier. The thickest ice ($4,522 \pm 10$ m) lies within the trough of the ASB, west of a second subglacial range, Highland B. In an assessment of bed data quality, the average difference in measured ice thicknesses where independently interpreted lines cross was found to be only 33 m.

Southeast portions of the ICECAP map compare well with the gross pattern found in previous compilations (including BEDMAP^{12,13}), which are largely constrained by airborne radar data from collaborative UK-US-Danish surveys from the 1970s¹⁴. Although a new Lagrangian interpolation¹⁵ of the sparse BEDMAP source data incorporating constraints from ice flow shows good general correspondence with the ICECAP data, direct assessment of the ICECAP profile data are required to understand the geomorphology of the region better. We use a conventional natural-neighbour interpolation¹⁶ in this paper.

In the northwest, ICECAP data indicate the presence of a deep depression inland from the Denman Glacier, confirming earlier results; however, instead of the 70,000-km² plateau suggested by the BEDMAP

Red lines are the profiles shown in Fig. 3. Major contours are 1,000 m apart.

b, Gridded along-track root mean squared deviation on an 800-m baseline. Bed elevation contours (500 m) are also marked. Profiles in Supplementary Figs 4 and 5 indicate the morphology of the rough highland regions. **c**, Three stages of ice-sheet development during the glacial fluctuations of the early Miocene or Oligocene epoch. Areas where the bed (without isostatic uplift) is above sea level are shown in yellow. Highland A constrains the edge of configuration 1; Highland B constrains the edge of configuration 2.

compilation, a smaller, intensely dissected mountain region (Highland A) lies to the west of the ASB (also see Supplementary Fig. 4).

The geomorphology uncovered by ICECAP data allows us to infer the nature of the former ice sheets in the region (Fig. 2a). About 20% of the ASB is more than 1 km below sea level, with the deepest regions located between 400 and 700 km inland from the present ice-sheet margin. Along-track bed roughness estimates (Fig. 2b) indicate that the ASB and the adjacent VSB are smooth (low vertical elevation changes over distances less than 1 km) when compared with their surroundings. The smooth domain of the ASB/VSB trough is not purely a function of elevation, but is bounded on its seaward side by a distinct ridge (Highland B) that is ~150 km wide, 0–500 m above sea level and deeply dissected by at least three ~50-km-wide valleys. Between Highland B and the ice-sheet margin, a broad, hummocky region hosts channels emanating from these valleys and heading towards the present-day margin. Highland A, to the west of the ASB, is also deeply dissected by a pair of parallel troughs each of which is more than 50 km wide. Transverse radar profiles reveal continuous troughs that are deepest where they traverse the highland regions (Fig. 3). These troughs reflect a high degree of morphological organization, given their similar size, shape and collinear positioning relative to one another.

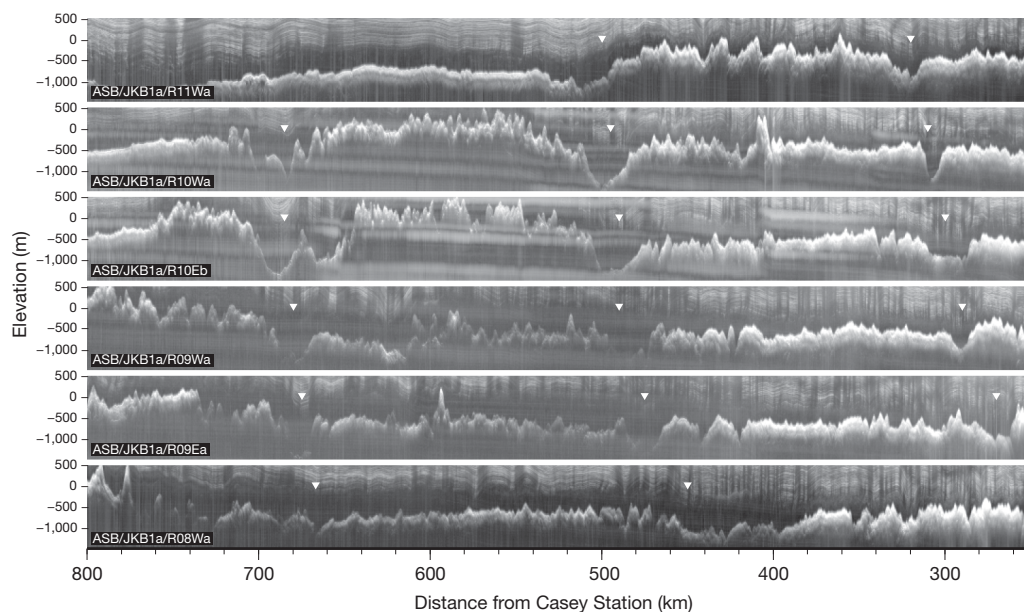


Figure 3 | Bed profiles of fjords in Highland B. Six depth-corrected radar profiles acquired using HiCARS radar are shown. These correspond to the red lines in Fig. 2a, and have similar orientation: south is to the left, west is to the top. Each adjacent profile extends from 500 m above to 1,500 m below sea level and shows a 550-km-long segment. Current sea level is at zero; isostatic uplift models (Supplementary Fig. 6) indicate that the bed may have been more than

500 m higher in the past. The major reflector in each profile is the bed reflection; above that lie layers within the ice. The ice surface is not shown. Fjords show pronounced over-deepening towards the ASB (in the upper half of the figure), which is reached at line R11Wa. Triangles indicate the axes of major through-cutting fjords.

It is clear that the topographic features are likely to be glacial in origin, but the nature of the ice masses responsible for their formation requires discussion. Ice sheets are most erosive near their margins, where high driving stresses, flow velocities and basal pressure gradients combine to produce distinct glacial geomorphology¹⁷. The modern ice sheet is unlikely to be capable of developing these features, because the regional ice flow is slow and cuts across trough axes (Supplementary Fig. 3). Hence, the morphology relates to past glaciations and to ice-sheet configurations different from those found today. Deep troughs selectively breaching uplands near the margins of ice sheets are well known in the Northern Hemisphere and occur in western Norway, East Greenland and eastern Baffin Island^{18–20}. Often such troughs have deepened pre-existing river valleys^{18–20}. In situations where such fjord troughs cut through the main upland axis, there is a close correlation between the depth of the trough and the height of the constraining uplands. Selective ice flow near an ice margin is favoured by two factors. First, the mountain barrier has a proportionately larger impact where the ice is thin and, second, ice velocities increase towards the ice-sheet margin. Under such conditions, low points in the topography are deepened and the more they deepen the more ice they drain^{21,22}, until an erosion threshold is reached¹⁷. In Antarctica, similar landscapes, such as the Transantarctic Mountains and the mountain front parallel to the coast in Dronning Maud Land, are associated with mountains acting as a barrier to ice flow and bounding the ice sheet.

Our understanding of the relationship between ice sheets and subglacial topography has benefited from analyses of formerly glaciated terrain and its glacial history, especially in Scandinavia. Here numerous episodes of glaciation centred on the main upland axis between Norway and Sweden have left a divergent pattern of regional troughs, and these have been overprinted by a continental-scale flow of ice across the region²³.

Two glacial configurations, in addition to the modern ice-sheet form, are inferred for the ASB. Configuration 1 involves ice flow from Ridge B that cuts deep channels in at least two places into Highland A to the west of the ASB. No ice from the Dome C region is needed to cut these valleys. A second series of valleys initiating in Highland B demonstrates ice flowing fast through the uplands east of the ASB

and across the broad, flat region towards the present-day margin. From this, we infer configuration 2, which is significantly larger in area than configuration 1 and involves convergent flow from Dome C and Ridge B into the ASB (Fig. 2c). The over-deepened troughs are formed through convergence of fast-flowing ice. They are manifest as topography shallowing downstream with reverse bed slope in the direction of the ice flow. Their formation requires an environment with abundant subglacial water¹⁷, probably requiring significant surface melt, analogous to Quaternary Northern Hemisphere glaciations (which are also noted for their oscillatory nature). The smooth landscape of the ASB upstream of fjords in configuration 2 is typical of a regime of enhanced erosion and deposition.

As in Scandinavia, we expect the valleys and troughs to show reactivation over time rather than to depict a single glacial event. In this way, the ASB may have experienced numerous glacial advances and recessions, many of which will have been orbitally paced. The glaciological reconstructions we infer from the landscape are consistent with the numerical models of growth and decay^{4,9,11}, despite the lack of detailed bed information in this sector informing these models. Ice growth and decay across the portions of the ASB that lie above sea level probably requires surface melt not currently present in Antarctica and, hence, temperatures significantly higher than at present. Such conditions, and therefore such ice sheets, have been restricted to the Northern Hemisphere over the past 14 Myr; hence, the formation of the landforms identified in the ASB highlands and, most probably, elsewhere in East Antarctica probably dates from the early Miocene or Oligocene epoch. If this is the case, the large oscillations in ice volume, paced by orbital changes, observed in offshore sequences²⁴ can be explained. An alternative view, that the glacial landforms were formed in the Pliocene epoch²⁴, requires the loss of much of the Antarctic ice sheet, with implications for global temperatures and sea levels.

Although it is difficult to know with certainty the topographic elevation of the ASB region during early EAIS oscillations, if the present ice sheet were removed the region seawards of the ice-cut fjords would be around sea level after isostatic uplift, whereas the ASB itself would remain substantially below sea level (Supplementary Fig. 6). As ice sheets are known to be sensitive to environmental change in such

lowland and shallow marine settings²⁵, we are able to infer the likely glacial processes responsible for changes in former ice sheets. Ice-sheet retreat to configuration 1 from configuration 2 may involve a marine instability similar to that proposed as being relevant to West Antarctica²⁶, in which ice retreat is associated with water depth increase at the margin and enhanced loss of ice through calving and melting leads to deglaciation of the entire ASB. Growth from configuration 1 to configuration 2 is more difficult to achieve, as it requires a major deep basin, filled with water, to be filled by grounded ice. Some have argued that deep, pre-glacial lakes such as Lake Vostok may have survived glaciation as subglacial lakes²⁷. Others have recognized the absence of large subglacial lakes in some troughs as evidence for migration of a grounded margin during ice growth²⁸. This is because a steep marginal surface gradient would drive water to the edge of the ice sheet. The absence of a large subglacial lake within the interior basin points to the latter explanation for its glaciation.

Evidence of fjords in East Antarctica cut by ice sheets of varying configuration may not be limited to our study region. Measurement of comparable features may allow us to appreciate better the magnitude of early EAIS change and the processes responsible.

METHODS SUMMARY

We used a ski-equipped, long-range DC-3T carrying a HiCARS coherent, 60-MHz, ice-penetrating radar²⁹ along with a gravimeter, magnetometers and laser altimeters. The out-and-back aircraft survey range is ~1,000 km. Twenty-six flights were supported by Casey Station in December–January of 2008–2009 and 2009–2010. Radial flights from Casey Station were undertaken to maximize coverage of the interior, along with reflights of ICESAT orbital tracks and coast-parallel tie lines. Radar data were pulse-compressed and processed using a short synthetic-aperture radar aperture to retain energy; with this level of processing, range distortions are not significant on length scales greater than 400 m. The ice thickness was found using a speed of light in ice of $169 \text{ m } \mu\text{s}^{-1}$, and the bed elevation was calculated using the radar-determined surface elevation. These new data were combined with data from BEDMAP and the Support Office for Aerogeophysical Research, and interpolated using a natural-neighbour algorithm¹⁶. Such algorithms are commonly used with irregularly distributed data confined to discrete transects. We determined along-track roughness using the root mean squared deviation³⁰ of detrended bed elevation data on an 800-m baseline.

Received 19 November 2010; accepted 12 April 2011.

1. Bo, S. *et al.* The Gamburtsev mountains and the origin and early evolution of the Antarctic Ice Sheet. *Nature* **459**, 690–693 (2009).
2. Zachos, J. C., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001).
3. Naish, T. R. *et al.* Orbitally induced oscillations in the East Antarctic ice sheet at the Oligocene/Miocene boundary. *Nature* **413**, 719–723 (2001).
4. DeConto, R. M. & Pollard, D. Rapid Cenozoic glaciation of Antarctica induced by declining atmospheric CO₂. *Nature* **421**, 245–249 (2003).
5. Pekar, S. F. & DeConto, R. M. High-resolution ice-volume estimates for the early Miocene: evidence for a dynamic ice sheet in Antarctica. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **231**, 101–109 (2006).
6. Summerfield, M. A. *et al.* Cosmogenic isotope data support previous evidence of extremely low rates of denudation in the Dry Valleys region, southern Victoria Land. *Spec. Publ. Geol. Soc. (Lond.)* **162**, 255–267 (1999).
7. Lewis, A. R., Marchant, D. R., Ashworth, A. C., Hemming, S. R. & Machlus, M. L. Major middle Miocene global climate change: evidence from East Antarctica and the Transantarctic Mountains. *Geol. Soc. Am. Bull.* **119**, 1449–1461 (2007).
8. Marchant, D. R. *et al.* Formation of patterned ground and sublimation till over Miocene glacier ice, southern Victoria Land, Antarctica. *Geol. Soc. Am. Bull.* **114**, 718–730 (2002).
9. Huybrechts, P. Glaciological modelling of the Late Cenozoic East Antarctic ice sheet: stability or dynamism? *Geogr. Ann.* **75**, 221–238 (1993).

10. Jamieson, S. S. R. & Sugden, D. E. in *Antarctica, a Keystone in a Changing World* (eds Cooper, A. *et al.*) 39–54 (National Academies, 2007).
11. Siegert, M. J., Taylor, J. & Payne, A. J. Spectral roughness of subglacial topography and implications for former ice-sheet dynamics in East Antarctica. *Global Planet. Change* **45**, 249–263 (2005).
12. Lythe, M., Vaughan, D. G. & the BEDMAP Consortium. BEDMAP: a new ice thickness and subglacial topographic model of Antarctica. *J. Geophys. Res.* **106**, 11335–11352 (2001).
13. Le Brocq, A. M., Payne, A. J. & Vieli, A. An improved Antarctic dataset for high resolution numerical ice sheet models (ALBMAP v1). *Earth Syst. Sci. Data* **2**, 247–260 (2010).
14. Drewry, D. J. Sedimentary basins of the East Antarctic craton from geophysical evidence. *Tectonophysics* **36**, 301–314 (1976).
15. Roberts, J. L. *et al.* Refined large-scale sub-glacial morphology of Aurora basin, East Antarctica derived by an ice-dynamics-based interpolation scheme. *Cryosphere Discuss.* **5**, 655–684 (2011).
16. Watson, D. *Contouring: A Guide to the Analysis and Display of Spatial Data* 67–68 (Pergamon, 1992).
17. Alley, R. B., Lawson, D. E., Larson, G. J., Evenson, E. B. & Baker, G. S. Stabilizing feedbacks in glacier-bed erosion. *Nature* **424**, 758–760 (2003).
18. Holtedahl, H. Notes on the formation of fjords and fjord valleys. *Geogr. Ann.* **49**, 188–203 (1967).
19. Sugden, D. E. Landscapes of glacial erosion in Greenland and their relationship to ice, topographic and bedrock conditions. *Inst. Br. Geogr. Spec. Publ.* **7**, 177–195 (1974).
20. Løken, O. H. & Hodgson, D. A. On the submarine geomorphology along the east coast of Baffin Island. *Can. J. Earth Sci.* **8**, 185–195 (1971).
21. Kessler, M. A., Anderson, R. S. & Briner, J. P. Fjord insertion into continental margins driven by topographic steering of ice. *Nature Geosci.* **1**, 365–369 (2008).
22. Jamieson, S. S. R., Hulton, N. R. J. & Hagdorn, M. Modelling landscape evolution under ice. *Geomorphology* **97**, 91–108 (2008).
23. Kleman, J., Stroeven, A. P. & Lundqvist, J. Patterns of Quaternary ice sheet erosion and deposition in Fennoscandia and a theoretical framework for explanation. *Geomorphology* **97**, 73–90 (2008).
24. Harwood, D. M., McMinn, A. & Quilty, P. G. Diatom biostratigraphy and age of the Pliocene Sørødal Formation, Vestfold Hills, East Antarctica. *Antarct. Sci.* **12**, 443–462 (2000).
25. Siegert, M. J. *Ice Sheets and Late Quaternary Environmental Change* 131–152 (Wiley, 2001).
26. Mercer, J. H. West Antarctic ice sheet and CO₂ greenhouse effect: a threat of disaster. *Nature* **271**, 321–325 (1978).
27. Duxbury, N. S., Zotikov, I. A., Nealson, K. H., Romanovsky, V. E. & Carsey, F. D. A numerical model for an alternative origin of Lake Vostok and its exobiological implications for Mars. *J. Geophys. Res.* **106**, 1453–1462 (2001).
28. Siegert, M. J. Comment on “A numerical model for an alternative origin of Lake Vostok and its exobiological implications for Mars” by N. S. Duxbury, I. A. Zotikov, K. H. Nealson, V. E. Romanovsky, and F. D. Carsey. *J. Geophys. Res.* **109**, E02007 (2004).
29. Peters, M. E., Blankenship, D. D. & Morse, D. L. Analysis techniques for coherent airborne radar sounding: application to West Antarctic ice streams. *J. Geophys. Res.* **110**, B06303 (2005).
30. Shepard, M. K. *et al.* The roughness of natural terrain: a planetary and remote sensing perspective. *J. Geophys. Res.* **106**, 32,777–32,795 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by NSF grant ANT-0733025 and NASA grant NNX09AR52G to the University of Texas at Austin, NERC grant NE/D003733/1 to the University of Edinburgh, Australian Antarctic Division project 3103, the Jackson School of Geoscience, and the Jet Propulsion Laboratory, and the G. Unger Vetlesen Foundation. This research was also supported by the Antarctic Climate and Ecosystems Cooperative Research Centre. This is UTIG contribution 2344.

Author Contributions D.A.Y., D.D.B., M.J.S., J.W.H., R.C.W., N.W.Y., J.L.R. and T.D.v.O. planned the investigation, including the flights. D.A.Y. and D.D.B. oversaw the data reduction. D.A.Y., D.D.B., A.P.W., J.W.H., J.S.G., D.M.S., J.L.R. and R.C.W. participated in the field work. D.E.S. and M.J.S. provided the geomorphic interpretation. D.A.Y., M.J.S., D.D.B. and A.P.W. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.A.Y. (duncan@utig.ig.utexas.edu) or M.J.S. (m.j.siegert@ed.ac.uk).

Strontium isotope evidence for landscape use by early hominins

Sandi R. Copeland^{1,2}, Matt Sponheimer², Darryl J. de Ruiter³, Julia A. Lee-Thorp^{4,5}, Daryl Codron⁶, Petrus J. le Roux⁵, Vaughan Grimes^{1,7} & Michael P. Richards^{1,8}

Ranging and residence patterns among early hominins have been indirectly inferred from morphology^{1,2}, stone-tool sourcing³, referential models^{4,5} and phylogenetic models^{6–8}. However, the highly uncertain nature of such reconstructions limits our understanding of early hominin ecology, biology, social structure and evolution. We investigated landscape use in *Australopithecus africanus* and *Paranthropus robustus* from the Sterkfontein and Swartkrans cave sites in South Africa using strontium isotope analysis, a method that can help to identify the geological substrate on which an animal lived during tooth mineralization. Here we show that a higher proportion of small hominins than large hominins had non-local strontium isotope compositions. Given the relatively high levels of sexual dimorphism in early hominins, the smaller teeth are likely to represent female individuals, thus indicating that females were more likely than males to disperse from their natal groups. This is similar to the dispersal pattern found in chimpanzees⁹, bonobos¹⁰ and many human groups¹¹, but dissimilar from that of most gorillas and other primates¹². The small proportion of demonstrably non-local large hominin individuals could indicate that male australopiths had relatively small home ranges, or that they preferred dolomitic landscapes.

Established palaeontological and archaeological techniques provide little tangible evidence of how early hominins used and moved across landscapes. For example, home-range size has been estimated on the basis of a rough correlation with body mass¹ and models of early hominin dispersal have relied on behaviours that are common among hominoids and are presumed to be present in a common ancestor^{6–8}.

We used a geochemical proxy, strontium isotope analysis of tooth enamel, to investigate landscape use in early hominins. We sampled a series of hominin tooth-crowns with a relatively new method for measuring strontium isotopes in teeth: laser ablation multi-collector inductively coupled plasma mass spectrometry (MC-ICP-MS). This method is almost non-destructive because it leaves only small traces (200 × 750 μm) on the enamel surface^{13,14} (Supplementary Fig. 1).

The ⁸⁷Sr/⁸⁶Sr ratios of animals directly reflect the ⁸⁷Sr/⁸⁶Sr ratios of their foods, which in turn are primarily a reflection of local geology and to a lesser extent, of hydrology and atmospheric inputs^{15–17}. The ⁸⁷Sr/⁸⁶Sr ratio of bedrock varies as a result of its original Rb/Sr ratio and age¹⁸. Because strontium is ingested and incorporated in trace quantities into mammalian teeth during the period of enamel mineralization, ⁸⁷Sr/⁸⁶Sr ratios can be used to study movement across landscapes in geologically heterogeneous areas. The fossil-hominin-bearing caves of the Sterkfontein Valley are situated in the diverse geological substrates of the Transvaal supergroup in South Africa, and represent one such area¹⁹. This presents a unique opportunity to investigate land-use patterns of South African early hominins from the Plio-Pleistocene.

We measured the ⁸⁷Sr/⁸⁶Sr ratios in the permanent dentition of 19 australopiths from the Sterkfontein Valley (Supplementary Table 1 and Methods). Because tooth-crown formation was complete by approximately age 8 in these hominins²⁰, the isotopic composition of enamel should reflect behaviour before dispersal from the natal group. The specimens included 11 *P. robustus* individuals from Swartkrans (Member 1, ~1.8 million years ago) and eight *A. africanus* individuals from Sterkfontein (Member 4, ~2.2 million years ago) (Supplementary Tables 1 and 2). For comparative purposes, we measured the ⁸⁷Sr/⁸⁶Sr ratios in the fossil enamel of 38 mammalian teeth that are roughly contemporaneous with these hominins (Supplementary Table 1). We use taxa with small home ranges (*Procavia* and *Raphicerus*), called here 'small-range mammals', as a proxy for local ⁸⁷Sr/⁸⁶Sr ratios²¹. To establish the biologically available ⁸⁷Sr/⁸⁶Sr ratios across the region, we analysed more than 170 modern plants and animals within 50 km of the fossil sites, sampling 11 different geological substrates (Fig. 1 and Supplementary Table 3).

The biologically available ⁸⁷Sr/⁸⁶Sr ratio of the geological substrate that includes the fossil-bearing caves, the Malmani dolomite, ranged from 0.721 to 0.734. Its mean differs from that of nine nearby geological substrates (nested analysis of variance on ranks, unequal N HSD, $P < 0.001$) and overlaps appreciably with that of only two substrates (Fig. 2 and Supplementary Table 3). Therefore, individuals with ⁸⁷Sr/⁸⁶Sr ratios outside the dolomite range can be identified with confidence as non-local to the dolomite, that is, the tooth was mineralized while the individual spent at least some time subsisting elsewhere. More specific spatial attribution, beyond non-local (non-dolomite), is not possible at this stage. The minimum distance from the Sterkfontein and Swartkrans caves to a non-local geology is 2–3 km to the southeast, 5–6 km to the northwest and >30 km in roughly northeast and southwest directions (Fig. 1).

The ⁸⁷Sr/⁸⁶Sr data for fossils indicated that a higher proportion of Hominidae (32%, $n = 19$) were non-local than the proportions of Cercopithecidae (14%, $n = 7$), Procaviidae (9%, $n = 11$) and Bovidae (16%, $n = 19$) (Supplementary Table 1 and Supplementary Discussion). Among 57 fossil teeth analysed, five out of seven of the teeth with the most radiogenic (high ⁸⁷Sr/⁸⁶Sr) values were hominins, and the ⁸⁷Sr/⁸⁶Sr ratios of the Hominidae were significantly different from those of the small-range mammals (Kruskal–Wallis, Conover–Inman, $P = 0.002$), whereas those of the other primate family, Cercopithecidae, were not.

Within the Hominidae, there was no statistically significant difference between the proportions of non-locals in *P. robustus* (36%) and *A. africanus* (25%). However, there were significant differences between subsets of hominins defined by tooth size (Supplementary Fig. 2, Supplementary Table 2 and Methods). The four most radiogenic specimens were small hominins (tooth size below the mean for that tooth and species) and at least 50% of the small hominins were non-local. In contrast, significantly fewer of the large hominins (11%; tooth size above the mean for that tooth and species) were

¹Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. ²Department of Anthropology, University of Colorado at Boulder, 233 UCB, Boulder, Colorado 80309, USA. ³Department of Anthropology, Texas A&M University, College Station, Texas 77843-4352, USA. ⁴Research Laboratory for Archaeology and the History of Art, Oxford University, Oxford OX1 3QY, UK. ⁵AEON EarthLAB, Department of Geological Sciences, University of Cape Town, Rondebosch 7701, South Africa. ⁶Clinic for Zoo Animals, Exotic Pets and Wildlife, Vetsuisse Faculty, University of Zurich, Winterthurerstrasse 260, 8057 Zurich, Switzerland. ⁷Department of Archaeology, Memorial University, St John's, Newfoundland A1C 5S7, Canada. ⁸Department of Anthropology, University of British Columbia, 6303 NW Marine Drive, Vancouver, British Columbia V6T 1Z1, Canada.

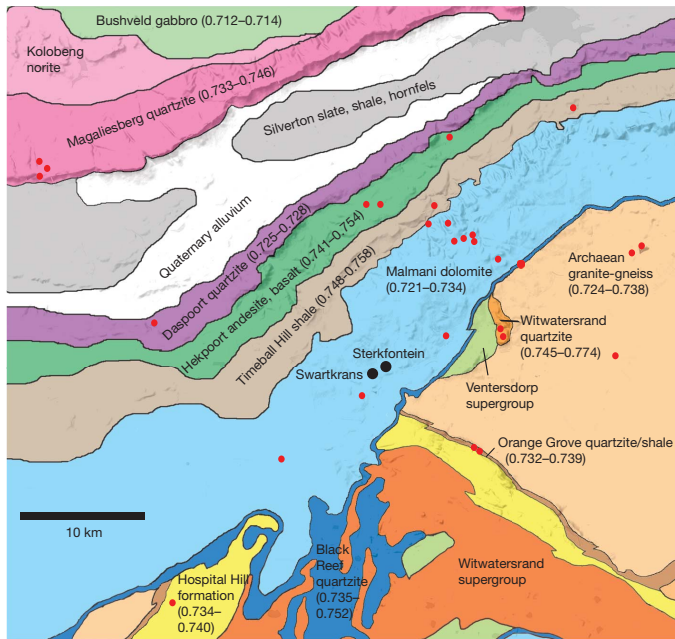


Figure 1 | Map of Sterkfontein Valley showing the locations of Sterkfontein and Swartkrans, geological zones and sampling areas. Geological zones are represented by different colours and their ranges of biologically available $^{87}\text{Sr}/^{86}\text{Sr}$ are labelled. Sampling localities for biologically available $^{87}\text{Sr}/^{86}\text{Sr}$ are shown as red circles. The figure is drawn on the basis of 1:250,000 maps published by the Geological Survey, Republic of South Africa, 1981. One sampling locality for Bushveld gabbro and two for Black Reef quartzite are not shown. Topographic relief in the area is mostly gentle and geological substrates adjacent to the Malmi dolomite are readily accessible on foot.

demonstrably non-local (Fisher's mid- P , $P = 0.049$) (Fig. 2 and Supplementary Table 1). Furthermore, there were more non-local small hominins than there were non-local small-range mammals (Fisher's mid- P , $P = 0.028$), whereas the proportions of non-local large hominins and non-local small-range mammals were indistinguishable.

Given the pattern of sexual dimorphism that has been suggested for *P. robustus* and *A. africanus*, both having male body weights of ~ 40 kg

and female body weights of ~ 30 kg (ref. 22), and the pattern of sexual dimorphism in tooth size among similarly dimorphic hominoids^{23,24}, it is tempting to refer to the larger and smaller teeth as belonging to males and females, respectively. However, because there is overlap in the tooth dimensions of male and female hominoids, sex assignments will be most secure for the largest and smallest individuals (see Supplementary Discussion). When we considered only teeth that were more than one standard deviation above or below the mean (Supplementary Fig. 2 and Supplementary Table 1), 75% of the smaller teeth ($n = 4$) were non-local and 17% of the larger teeth ($n = 6$) were non-local, upholding the pattern found for the entire sample.

Thus, the strontium isotope data indicate differences in landscape use between males and females. Because Sr was incorporated into the teeth before adulthood, when the hominins were probably travelling with their mothers, the data are unlikely to reflect differences in foraging areas between adult males and adult females. Rather, the Sr isotopes probably indicate that females preferentially moved away from residential groups. Models of the earliest hominins, or of the last common ancestor of African apes and hominins, have often proposed female exogamy^{6–8} because, although it is rare among primates in general, it is a trait shared by some modern human groups¹¹, *Pan*^{9,10} and *Gorilla*¹². *Pan* and many modern humans are characterized by male philopatry^{9–11}, whereas in *Gorilla*, both males and females disperse at similar rates¹².

A *Gorilla*-like social structure was proposed for *P. robustus* on the basis of craniodental evidence for extended male development²⁵. In *Gorilla*, one or a few males monopolize groups of females, forcing some males to disperse from their natal groups¹². Therefore, a *Gorilla*-like social structure might be expected to produce similar evidence for dispersal among male and female hominins. The strontium isotope data for *P. robustus* are too few to address this question, but differential evidence for male and female dispersal among the South African australopiths generally is more consistent with *Pan*-like dispersal patterns (see Supplementary Discussion). The lower sexual dimorphism in the size of canine teeth, relative to *Pan* and *Gorilla*²⁴, may also militate against both *Gorilla*-like and *Pan*-like social structures because it indicates relatively low competition between males. We think it is likely that there is no appropriate modern analogue for the social structure of these australopiths, given their marked anatomical and ecological differences from extant hominoids (see ref. 26).

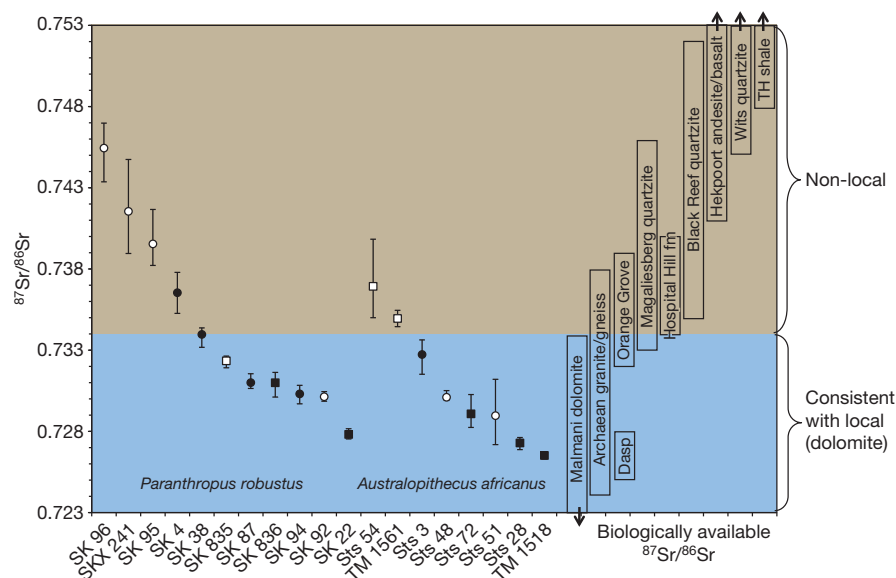


Figure 2 | Strontium isotope ratios of australopith tooth enamel and biologically available $^{87}\text{Sr}/^{86}\text{Sr}$ ratios across the Sterkfontein Valley. Symbols represent the mean $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of multiple laser scans made on each tooth (white, small tooth; black, large tooth; circle, canine; square, third molar). Whiskers show the intra-tooth range of laser scans. Internal (analytical) error

for each laser scan was <0.0003 (2σ). Biologically available $^{87}\text{Sr}/^{86}\text{Sr}$ ratios are based on the minimum to maximum $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of plants collected on each geological substrate. Ranges for Malmi dolomite, Hekpoort andesite/basalt, Witwatersrand (Wits) quartzite, Timeball Hill (TH) shale and Bushveld gabbro extend beyond the chart scale. Daspoort quartzite; fm, formation.

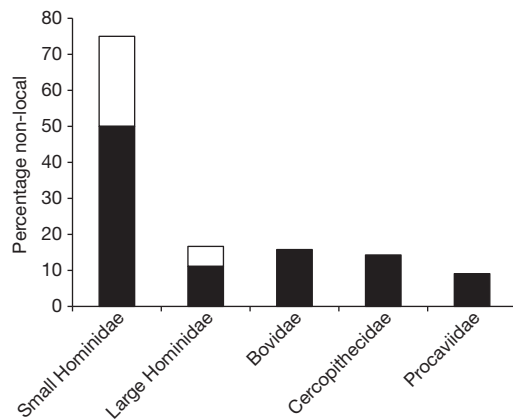


Figure 3 | Proportions of non-local individuals among the fossil specimens. Bars represent proportions of fossil specimens for which the $^{87}\text{Sr}/^{86}\text{Sr}$ ratio falls outside the range of biologically available $^{87}\text{Sr}/^{86}\text{Sr}$ for the Malmani dolomite. The black bars include all non-local specimens of each group; black bars plus white areas indicate the proportions of non-local individuals when only hominin specimens less than or more than one standard deviation from mean tooth size for that category of tooth are considered (Supplementary Fig. 2 and Supplementary Discussion).

The higher proportion of non-local Hominidae compared to other taxa (Fig. 3) does not necessarily indicate that all hominins ranged more widely; only that small individuals did, and we hypothesize that this is best explained by dispersal events. The small proportion of non-local large australopithecids provides little evidence that males of either species ranged off the dolomite routinely. This could imply that these hominins had relatively small annual home ranges and/or that they had a preference for habitats on the dolomites (for example, owing to the presence of caves or preferred vegetation, see Supplementary Discussion), with potential implications for our understanding of locomotor adaptations²⁷, body proportions²⁸ and dietary ecology in early hominins.

METHODS SUMMARY

All fossil teeth analysed are housed at the Ditsong National Museum of Natural History (formerly the Transvaal Museum) in Pretoria. Hominin tooth size was determined by comparing each tooth specimen to a sample that included all measurable specimens of each tooth category recovered from Sterkfontein, Makapansgat, Kromdraai, Swartkrans and Drimolen (see Supplementary Fig. 2).

All of the fossil hominin teeth and all but two of the non-hominin teeth (see Supplementary Table 1) were analysed by laser ablation MC-ICP-MS and were sampled with several linear laser scans located on the external surface of the enamel. We spaced these laser scans evenly along the growth axis of the tooth, with the long axis of each linear laser scan ($200 \times 750 \mu\text{m}$) oriented perpendicular to the growth axis of the tooth. Laser ablation MC-ICP-MS was performed using a New Wave UP213 laser ablation unit operating at 213 nm, coupled to a NuPlasma high-resolution MC-ICP-MS in the Africa Earth Observatory Network (AEON) EarthLAB Facility at the University of Cape Town, following methods outlined in ref. 14. The quality of the data was checked by monitoring a bracketing standard (a modern rodent tooth) before and after all laser scans on fossil specimens during each laser session (Supplementary Fig. 1 and Supplementary Table 4).

Plant specimens for determining biologically available $^{87}\text{Sr}/^{86}\text{Sr}$ ratios were collected within a 50 km radius of the Sterkfontein and Swartkrans caves in areas where the soil was not contaminated by agriculture, mining, etc. Plants were dried and reduced to ash in the Department of Archaeology at the University of Cape Town. The plant ash was analysed for $^{87}\text{Sr}/^{86}\text{Sr}$ by solution MC-ICP-MS at the Max Planck Institute for Evolutionary Anthropology in Leipzig.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 December 2010; accepted 14 April 2011.

1. Antón, S. C., Leonard, W. R. & Robertson, M. An ecomorphological model of the initial hominid dispersal from Africa. *J. Hum. Evol.* **43**, 773–785 (2002).
2. Wheeler, P. E. The thermoregulatory advantages of large body size for hominids foraging in savanna environments. *J. Hum. Evol.* **23**, 351–362 (1992).

3. Ambrose, S. H. Paleolithic technology and human evolution. *Science* **291**, 1748–1753 (2001).
4. McGrew, W. C. in *Woman the Gatherer* (ed. Dahlberg, F.) 35–73 (Yale Univ. Press, 1981).
5. Moore, J. in *Great Ape Societies* (eds McGrew, W. C., Marchant, L. F. & Nishida, T.) 275–292 (Cambridge Univ. Press, 1996).
6. Ghiglieri, M. P. Sociobiology of the great apes and the hominid ancestor. *J. Hum. Evol.* **16**, 319–357 (1987).
7. Wrangham, R. W. in *The Evolution of Human Behavior: Primate Models* (ed. Kinzey, W. G.) 51–71 (State Univ. of New York Press, 1987).
8. Foley, R. & Gamble, C. The ecology of social transitions in human evolution. *Phil. Trans. R. Soc. B* **364**, 3267–3279 (2009).
9. Goodall, J. *The Chimpanzees of Gombe: Patterns of Behavior* (Harvard Univ. Press, 1986).
10. Furuichi, T. Social interactions and the life history of female *Pan paniscus* in Wamba, Zaire. *Int. J. Primatol.* **10**, 173–197 (1989).
11. Ember, C. R. Myths about hunter-gatherers. *Ethnology* **17**, 439–448 (1978).
12. Robbins, M. M. et al. Social structure and life-history patterns in western gorillas (*Gorilla gorilla gorilla*). *Am. J. Primatol.* **64**, 145–159 (2004).
13. Copeland, S. R. et al. Strontium isotope ratios in fossil teeth from South Africa: assessing laser ablation MC-ICP-MS analysis and the extent of diagenesis. *J. Archaeol. Sci.* **37**, 1437–1446 (2010).
14. Copeland, S. R. et al. Strontium isotope ratios ($^{87}\text{Sr}/^{86}\text{Sr}$) of tooth enamel: a comparison of solution and laser ablation MC-ICP-MS methods. *Rapid Commun. Mass Spectrom.* **22**, 3187–3194 (2008).
15. Hurst, R. W. & Davis, T. E. Strontium isotopes as tracers of airborne fly ash from coal-fired power plants. *Environ. Geol.* **3**, 363–367 (1981).
16. Gosz, J. R., Brookins, D. G. & Moore, D. I. Using strontium isotope ratios to estimate inputs to ecosystems. *Bioscience* **33**, 23–30 (1983).
17. Graustein, W. C. in *Stable Isotopes in Ecological Research* (eds Rundel, P. W., Ehleringer, J. R. & Nagy, K. A.) 491–512 (Springer, 1989).
18. Faure, G. & Powell, T. *Strontium Isotope Geology* (Springer, 1972).
19. Sillen, A., Hall, G., Richardson, S. & Armstrong, R. $^{87}\text{Sr}/^{86}\text{Sr}$ ratios in modern and fossil food-webs of the Sterkfontein Valley: implications for early hominid habitat preferences. *Geochim. Cosmochim. Acta* **62**, 2463–2473 (1998).
20. Beynon, A. D. & Dean, M. C. Distinct dental development patterns in early fossil hominids. *Nature* **335**, 509–514 (1988).
21. Price, T. D., Burton, J. H. & Bentley, R. A. The characterization of biologically available strontium isotope ratios for the study of prehistoric migration. *Archaeometry* **44**, 117–135 (2002).
22. McHenry, H. M. & Coffing, K. *Australopithecus to Homo: transformations in body and mind*. *Annu. Rev. Anthropol.* **29**, 125–146 (2000).
23. Mahler, P. E. *Metric Variation in the Pongid Dentition*. PhD thesis, Univ. Michigan (1973).
24. Plavcan, J. M. & van Schaik, C. P. Interpreting hominid behavior on the basis of sexual dimorphism. *J. Hum. Evol.* **32**, 345–374 (1997).
25. Lockwood, C. A., Menter, C. G., Moggi-Cecchi, J. & Keyser, A. W. Extended male growth in a fossil hominin species. *Science* **318**, 1443–1446 (2007).
26. Tooby, J. & DeVore, I. in *The Evolution of Human Behavior: Primate Models* (ed. Kinzey, W. G.) 183–237 (State Univ. of New York Press, 1987).
27. Rodman, P. S. & McHenry, H. M. Bioenergetics and the origin of hominid bipedalism. *Am. J. Phys. Anthropol.* **52**, 103–106 (1980).
28. McHenry, H. M. & Berger, L. Body proportions in *Australopithecus afarensis* and *A. africanus* and the origin of the genus *Homo*. *J. Hum. Evol.* **35**, 1–22 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Funding for this project was provided by the National Science Foundation, USA (grant 0609963), the Max Planck Society, a University of Colorado LEAP Associate Professor Growth Grant and the University of Colorado Dean's Fund for Excellence. We are grateful to the Ditsong National Museum of Natural History (formerly the Transvaal Museum), including F. Thackeray, S. Potze and T. Kearny, for allowing us access to the Sterkfontein and Swartkrans fossil material and for granting permission for laser ablation MC-ICP-MS analysis. We thank S. Potze and T. Perregil for flying to Cape Town to allow laser ablation MC-ICP-MS sampling of the hominins. We also thank L. Berger, J. Brophy, J. Codron and J. Hancox for help in the field, J. Montgomery for useful discussions and J. Sealy and J. Lanham for help with plant preparation in the Archaeology Department, University of Cape Town. We thank A. Anderson, C. Campbell, B. Covert, D. Grimstead, F. Grine, J. Leichter, O. Paine, J. Quade, P. Sandberg and P. Ungar for commenting on the manuscript. The research of A. Sillen and G. Hall inspired this study.

Author Contributions M.S. and J.A.L.-T. conceived the project. S.R.C., M.S., D.J.d.R., J.A.L.-T. and D.C. conducted fieldwork. D.J.d.R. chose hominin tooth specimens and made occlusal measurements. S.R.C., D.J.d.R., J.A.L.-T. and P.J.I.R. performed laser ablation MC-ICP-MS analyses. S.R.C. and V.G. performed solution MC-ICP-MS analyses. M.P.R. directed analyses at MPI-EVA. S.R.C., M.S., D.J.d.R. and D.C. wrote the manuscript. All authors discussed the results and commented on the manuscript. M.S. was principal investigator for the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.R.C. (sandi.copeland@colorado.edu).

METHODS

All fossil teeth analysed are housed at the Ditsong National Museum of Natural History (formerly the Transvaal Museum) in Pretoria. Hominin tooth size was determined by comparing each tooth specimen to a sample that included all measurable specimens of each tooth category recovered from Sterkfontein, Makapansgat, Kromdraai, Swartkrans and Drimolen (see Supplementary Fig. 2 and Supplementary Discussion).

Fossil teeth were analysed by laser ablation MC-ICP-MS, with the exception of two *Antidorcas marsupialis* teeth, small pieces of which were analysed by solution MC-ICP-MS (Supplementary Table 1). All of the hominin teeth and most of the faunal teeth were sampled with multiple linear laser scans (Supplementary Table 1) located on the external surface of the enamel, but not on the occlusal surface. The area was prepared with successive swabbing and drying episodes using ultrapure deionized water, acetone and 0.1 M acetic acid.

Wherever possible, the linear laser scans ($200 \times 750 \mu\text{m}$ per scan for the analysis run) were evenly spaced along the growth axis of the tooth, with the long axis of each laser scan oriented perpendicular to the growth axis of the tooth. The distance between laser scans on hominin teeth ranged from 0.4 to 2.4 mm and the height of crown available for sampling was 5.2–11.7 mm. The number of laser scans per tooth did not correlate with the variability within a tooth (with variability defined as the maximum laser-scan $^{87}\text{Sr}/^{86}\text{Sr}$ ratio minus the minimum laser-scan $^{87}\text{Sr}/^{86}\text{Sr}$ ratio; $R^2 = 0.018$), nor did the length of crown measured correlate with variability ($R^2 = 0.117$).

Laser ablation MC-ICP-MS was performed using a New Wave UP213 laser ablation unit operating at 213 nm, coupled to a NuPlasma high-resolution MC-ICP-MS in the Africa Earth Observatory Network (AEON) EarthLAB Facility at the University of Cape Town, following the methods detailed in ref. 14 (laser set-up parameters are given in Supplementary Table 5).

The quality of the data was checked by monitoring a bracketing standard before and after all laser scans on specimens during each laser session (Supplementary Fig. 1 and Supplementary Table 4). The standard was a modern rodent incisor (either specimen 26-r1 or specimen 26-10z, depending on availability; Supplementary Table 4) from a modern barn owl roost at Gladysvale cave (about 13 km northeast of Sterkfontein). 'True' $^{87}\text{Sr}/^{86}\text{Sr}$ ratios for the standards were determined by solution MC-ICP-MS, with solution data referenced to the international Sr isotope standard, SRM 987 ($^{87}\text{Sr}/^{86}\text{Sr} = 0.710255$). Strontium isotope ratios of the fossil specimens were not corrected to the bracketing standard; rather,

results were only accepted from laser sessions in which the $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of the bracketing standard were within 0.001 of their solution values.

$^{87}\text{Sr}/^{86}\text{Sr}$ ratios of fossil enamel in two specimens (SKX 8455a and SKX 4878, both *Antidorcas marsupialis*) were measured by solution MC-ICP-MS. A ~5 mg piece of fossil enamel was removed from the tooth in South Africa using a handheld drill outfitted with a diamond-bit saw and was then taken to the Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology (MPI-EVA) for analysis, following the methods outlined in ref. 14 (operation parameters are given in Supplementary Table 6).

Plant specimens for determining biologically available $^{87}\text{Sr}/^{86}\text{Sr}$ ratios were collected within a 50 km radius of the Sterkfontein and Swartkrans cave sites. For each geological substrate, up to ten sampling sites were selected in areas away from agricultural land-use, and in areas not disturbed by road development, mining or other activities that might contaminate the soil with foreign substances. At each sampling site, leaves, stems, seeds and/or fruits were collected from up to ten different plants, which included a variety of grass, forb and tree species. For each plant, about 2 g of air-dried plant material was reduced to ash in a porcelain crucible in a muffle oven at 500 °C for 6–8 h in the Department of Archaeology, University of Cape Town.

In the clean lab at MPI-EVA, 10–20 mg of plant ash was weighed into Teflon or Savillex beakers and digested in 2 ml of 14.3 M (65%) HNO_3 at 120 °C for 8 h. The remains were centrifuged three times at 13,000 r.p.m. for 20 min. After separation, the clear liquid was dried down and re-dissolved in 1 ml of 3 M HNO_3 . The sample was then loaded onto clean 2-ml columns containing preconditioned²⁹ 50–100 μm bead Sr-Spec resin. After three washes with 3 M HNO_3 , the Sr was eluted in ultrapure deionized water and dried, then re-dissolved in 3% HNO_3 in preparation for solution MC-ICP-MS.

Plant and tooth samples for solution MC-ICP-MS were analysed at MPI-EVA on a Thermo Fisher Neptune MC-ICP-MS instrument in static mode, following the methods detailed in ref. 14 (operation parameters are in Supplementary Table 6). Repeated measurements of strontium isotope standard SRM 987, which was run concurrently with the samples, resulted in a mean $^{87}\text{Sr}/^{86}\text{Sr}$ value of 0.710264 ± 0.000035 (2σ , $n = 66$). Total procedural blanks were considered negligible ($^{88}\text{Sr} = 0.040\text{--}0.070 \text{ V}$) at <0.4% of the analyte signal intensity ($^{88}\text{Sr} \sim 20 \text{ V}$).

29. Charlier, B. L. A. *et al.* Methods for the microsampling and high-precision analysis of strontium and rubidium isotopes at single crystal scale for petrological and geochronological applications. *Chem. Geol.* **232**, 114–133 (2006).

Nematoda from the terrestrial deep subsurface of South Africa

G. Borgonie¹, A. García-Moyano^{2†}, D. Litthauer², W. Bert^{1,3}, A. Bester², E. van Heerden², C. Möller², M. Erasmus² & T. C. Onstott⁴

Since its discovery over two decades ago, the deep subsurface biosphere has been considered to be the realm of single-cell organisms, extending over three kilometres into the Earth's crust and comprising a significant fraction of the global biosphere^{1–4}. The constraints of temperature, energy, dioxygen and space seemed to preclude the possibility of more-complex, multicellular organisms from surviving at these depths. Here we report species of the phylum Nematoda that have been detected in or recovered from 0.9–3.6-kilometre-deep fracture water in the deep mines of South Africa but have not been detected in the mining water. These subsurface nematodes, including a new species, *Halicephalobus mephisto*, tolerate high temperature, reproduce asexually and preferentially feed upon subsurface bacteria. Carbon-14 data indicate that the fracture water in which the nematodes reside is 3,000–12,000-year-old palaeometeoritic water. Our data suggest that nematodes should be found in other deep hypoxic settings where temperature permits, and that they may control the microbial population density by grazing on fracture surface biofilm patches. Our results expand the known metazoan biosphere and demonstrate that deep ecosystems are more complex than previously accepted. The discovery of multicellular life in the deep subsurface of the Earth also has important implications for the search for subsurface life on other planets in our Solar System.

Phylum Nematoda Potts, 1932
Suborder Cephalobina
Superfamily Panagrolaimidae
Halicephalobus mephisto sp. nov.

Etymology. *Mephisto* (from *Mephistopheles*, pseudo-Greek): “he who loves not the light”, alluding to the Devil, Lord of the Underworld, in reference to the Faust legend in medieval mythology because the new species is found at a depth of 1.3 km in the Earth's crust.

Holotype. Museum voor Dierkunde, Ghent University, Belgium (collection number, UGMD 104182).

Paratype. Museum voor Dierkunde, Ghent University, Belgium (collection number, UGMD 104182) (nine paratypes); University of the Free State, Bloemfontein, South Africa (collection number, UFS GB0035) (six paratypes).

Locality. Type population collected from shaft 3, level 26, corridor 28 of Beatrix gold mine, South Africa, at a depth of 1.3 km, approximately 1 km north of shaft 3 (28° 14' 24.06'' S, 26° 47' 45.25'' E). Nematodes collected from fracture water expelled from a high-pressure valve.

Diagnosis. Body straight to slightly ventrally arcuate after fixation, 0.52–0.56 mm long with annulations. Tail relatively long and tail tip filiform; terminus straight to variably curved. Reproductive system monodelphic, prodelphic and on right side of intestine with posterior reflexed ovary extending 99–135 µm posterior to vulva. Ovary tip not reflexed back anteriorly. Temperature tolerant and parthenogenetic. See Supplementary Discussion and Description for more details.

Description. Although *Halicephalobus* is a morphologically minimalist genus, *H. mephisto* is a new typological morphospecies as it can

be easily differentiated from all other species of *Halicephalobus* by the presence of a long tail (110–130 µm; ratio of total length to body diameter at anus, $c' = 9$ –10) with a filiform terminus and no reflexed ovary tip (Fig. 1). Phylogenetically, *H. mephisto* has a maximally supported sister relationship with *Halicephalobus gingivalis*–*Halicephalobus* spp., differing by 10% from *H. gingivalis* and by 8% from other *Halicephalobus* species. In comparison, the most closely related different genus, *Procephalobus*, differs by 17% from *H. mephisto*. Alternative alignment methods did not have a single effect on the tree topology outcome and resulted in a similar or increased branch support for the *H. gingivalis*–*Halicephalobus* spp. clade. The monophyly of *Halicephalobus* was always maximally supported, independent of the alignment method. Furthermore, 13 and 6 autapomorphic characters were present in the small-subunit ribosomal RNA sequences for the new species and its sister clade, respectively. Although based on limited available homologous sequences, multiple autapomorphic characters from two loci (18S and D2D3) and multiple autapomorphic characters in both sister lineages (18S) indicate lineage exclusivity for *H. mephisto* with respect to other *Halicephalobus* species (Fig. 2). Hence, the species status of *H. mephisto* fulfils the requirements of an amalgamation of evolutionary and phylogenetic species concepts.

Although Eukaryota, Bacteria and Archaea cohabitate in almost all surface environments on the Earth, very few searches for eukaryotes in the subsurface have been published. In South Carolina 0.1–10 eukaryotes per gram comprising algae, fungi, amoebae and flagellates have been discovered⁵ at a depth of 200 m, and in Sweden 0.01–1 fungal cells per millilitre, ~3 µm in size, have been found⁶ in 200–450-m deep fractures. In this study of the South African subsurface, we expanded the search for subsurface life to nematodes because they are one of the most successful metazoan phyla with respect to their abundances, distribution and physiological tolerance^{7,8}; they can enter a state of anabiosis for extended periods; and they continue to metabolize aerobically in hypoxic environments where the partial pressure of oxygen (p_{O_2}) is only 0.4 kPa (ref. 9).

We took the following steps to determine whether the nematodes recovered were indigenous and not recent surface or mining contaminants: adaptation of filtration procedures that had been successfully used to collect planktonic microorganisms from thousands of litres of borehole water¹⁰ (Supplementary Methods); soil samples around the boreholes and the mining water were tested for nematodes; the chemical composition and the microbial community structure of the fracture water were determined; and the ³H and ¹⁴C concentrations were measured. Twenty-two water samples were collected from six boreholes ranging in depth from 0.5 to 3.6 km and located in five different South African mines (Table 1). Between 475 and 9,792 l of water were filtered for each sample. Eighteen soil samples were collected, six of which were from Beatrix gold mine. Seven mining water samples were collected, three of which were from Beatrix gold mine and for which 2 to 31,582 l of water were filtered.

¹Department of Biology, Nematology Section, Ghent University, Ledeganckstraat 35, B9000 Ghent, Belgium. ²Metagenomics Platform, Department of Microbial, Biochemical and Food Biotechnology, University of the Free State, PO Box 339, Bloemfontein 9300, South Africa. ³Laboratory of Nematology, Department of Plant Sciences, Wageningen University, 6708 Wageningen, The Netherlands.

⁴Department of Geosciences, Princeton University, Princeton, New Jersey 08544, USA. †Present address: Department of Biology, University of Bergen, Postbox 7803, N-5020 Bergen, Norway.

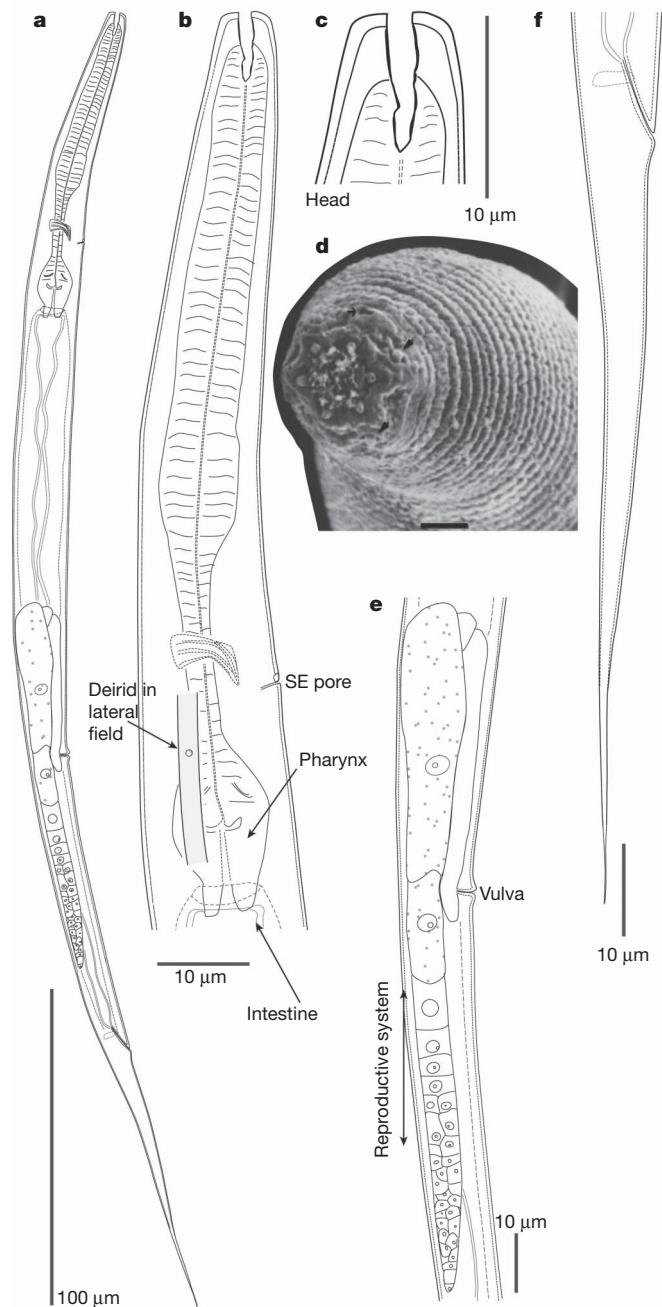


Figure 1 | General morphology of *H. mephisto*. Light microscopy drawings of female holotype and scanning electron microscopy photograph of head. **a**, Entire body; **b**, neck region; **c**, anterior region; **d**, scanning electron microscope face view (scale bar, 1 µm; black arrowheads indicate the positions of two cephalic papillae; black arrow indicates amphid opening); **e**, reproductive system; **f**, tail. SE, secretory–excretory.

Borehole water from Beatrix gold mine yielded the new species, *H. mephisto*. Borehole water from Driefontein gold mine yielded two nematode species, *Plectus aquatilis* and a monhysterid specimen that survived but did not reproduce. Borehole water from the shallowest site, Star Diamonds mine, and from Zondereinde platinum mine did not yield any nematodes, but a fourth nematode, a monhysterid species, was detected in DNA extracted from borehole water from the deepest site at Tau Tona gold mine. All three living nematode species fed on the borehole water bacteria in preference to *Escherichia coli* (Supplementary Methods), and the two that were able to reproduce did so by parthenogenesis (Table 2). None of these species were found in more than one borehole. Although these three boreholes were sampled on

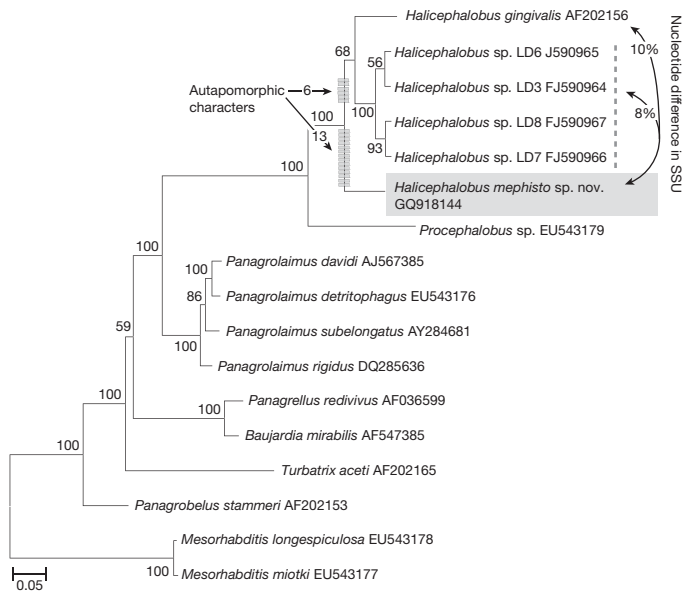


Figure 2 | Bayesian-interference 50%-majority-rule consensus phylogenies based on small-subunit rDNA data. *H. mephisto* with GenBank sequences of closely related taxa. Branch support is indicated with posterior probability values. Scale bar, expected substitutions per site. SSU, small subunit.

multiple dates, nematodes were only recovered from the first sample (Table 1). No other metazoans were detected. Only two of the soil samples yielded nematodes, but only when the soil was wet, and all were taxonomically distinct from the three borehole nematodes (Table 1). None of the soil samples from Beatrix gold mine yielded nematodes. Nematodes also were not detected in any mining water samples.

The borehole valves at the Beatrix and Driefontein gold mines were closed for at least several months to a year before sampling and the high water pressure on opening the valves and flushing the filtration apparatus precluded contamination from air. The possibility of nematodes being contaminants from borehole drilling is unlikely because the mining water used for drilling is treated with NaClO and H₂O₂ to the point that nematode DNA is highly degraded and because the high pressure of fracture water encountered during drilling flushes out drilling water (Supplementary Discussion). Nonetheless, we tested for this possibility at Beatrix gold mine—where a 6,480-l borehole water sample yielded *H. mephisto*—by filtering 31,582 l of the mining water used for drilling from a valve close to the borehole: no nematodes were detected. We cannot preclude the possibility that nematodes were present in the mining water at the time the borehole was drilled a year before sampling, but this seems unlikely given that the disinfection procedures are standard operating protocols. Because the nematodes were parthenogenetic, crossing with surface species was impossible, and nematode morphology and genetics cannot be used as indicators of long-term isolation^{11,12}. Further evaluation of their indigeneity, therefore, relied upon environmental data.

The water sampled was hypoxic, with dissolved O₂ concentrations ranging from 13 to 72 µM (Supplementary Tables). The borehole water from Beatrix gold mine yielded sulphate concentrations that were similar to that of the 3–5-Myr-old fracture water from a depth of 1.5 km (ref. 13) and were 100 times less than that of the mining water¹⁴. The geochemistry of the borehole water from Driefontein gold mine was consistent with groundwater from the karstic, sulphidic Transvaal dolomite aquifer¹⁴. The borehole water from Tau Tona gold mine was consistent with other highly saline fracture water from depths of >3 km and was distinct from the mining water and acidic mine drainage water in this mining district^{14,15}. The 16S rRNA gene clone libraries comprised sulphate-reducing Firmicutes and Deltaproteobacteria, heterotrophic Proteobacteria (fermenters and/or methanotrophs), Nitrospira and chemolithotrophic Proteobacteria that have

Table 1 | Geochemical, isotopic and nematode results

	Star Diamonds	Driefontein	Beatrix		Zondereinde	Tau Tona
			Borehole 1	Borehole 2		
Depth (km)	0.5	0.9	1.3	1.3	1.7	3.6
T (°C)	32	24	30	37	48	48
pH	8.3	7.5	7.7	7.9	8.3	7.7
pO ₂ (kPa)*	4.6	2.5	3.1	1.3–6.8	2.4	2.9
Microbial counts (cells per litre)	NA	<10 ⁵	NA	3 × 10 ⁶	2 × 10 ⁵	3.4 × 10 ⁶
Samples with nematodes/total samples	0/2	1/3†	0/6	1/6	0/1	1/4
Nematodes (number per litre)	NA	~7 × 10 ⁻⁵	NA	~3 × 10 ⁻⁵	<10 ⁻⁴	>~5 × 10 ⁻⁴
³ H (TR)	NA	0.270 ± 0.026	NA	0.014 ± 0.024	0.062 ± 0.045	0.034 ± 0.021
δ ¹³ C (‰ VPDB)‡	NA	-8	NA	-32	-17.4	-17.7
Δ ¹⁴ C§	NA	-932.8 ± 1.0	NA	-704.1 ± 2.2	-645.9 ± 2.4	-619.6 ± 2.9
¹⁴ C age (yr)	NA	10,104–12,084	NA	4,413–6,247	5,798	2,919–5,165
<i>H. mephisto</i>	—	—	—	+	—	—
<i>P. aquatilis</i>	—	+	—	—	—	—
Monhysterid sp. 1	—	+	—	—	—	—
Monhysterid sp. 2	—	—	—	—	—	+
Control samples						
Soil¶	Dorylaimids, mononchids and annelids	<i>Diploscapter coronatus</i> and <i>Rhabditis regenfussi</i>	—	—	—	—
Mining water#	—	—	—	—	—	—

NA, not analysed; TR, tritium units (1 ³H per 10¹⁸ H atoms).

*pO₂ (kPa) = 101.325 × [O₂] × 1.8 × 10⁴/K_H, where K_H = exp{[-286.942 + 15,450.6/T(K) + 36.5593ln(T(K)) + 0.0187662T(K)]/1.987} is Henry's solubility constant (moles of O₂ per moles of H₂O at the partial pressure of O₂ in atmosphere) and [O₂] is the dissolved O₂ concentration in micromolar.

†One sample yielded two nematodes of two separate species. In the case of Beatrix gold mine, only one nematode was found.

‡δ¹³C = (13C/12C)_{sample}/(13C/12C)_{standard} - 1; standard used is Vienna Pee Dee *Belemnite americana* (VPDB).

§Δ¹⁴C = 10 × 14C_{PMC} - 1,000; 14C_{PMC} is the carbon activity as per cent modern carbon.

|| Samples did not yield a cultivable nematode, but the DNA from the fracture yielded an 18S rRNA gene sequence belonging to Monhysteridae.

¶ Approximately 900 g of soil was plated, and in the cases where the soil was dry an additional 900 g was wetted and plated. The total number of nematodes was 17 for Star Diamonds mine and 18 for Driefontein gold mine.

Two to six litres of mining water was filtered for metazoans except at Beatrix gold mine, where 31,582 litres of mining water was filtered (Supplementary Tables and Supplementary Discussion).

been previously detected in the fracture water of these mines¹⁵ (Supplementary Tables and Supplementary Discussion). With the exception of the Driefontein borehole water, the ³H concentrations were within 2 s.d. of the detection limit (Table 1). In comparison with the concentration, 10–100 TR, in regional precipitation during the late 1980s¹⁶, these values indicate that <1% of the borehole water comprises post-1980s surface water and that no more than ~3% of the Driefontein borehole water can be modern. The Δ¹⁴C values for the dissolved inorganic carbon ranged from -932.8 to -619.6 and, using these values along with regional recharge values and corrections for dead carbon (Supplementary Tables and Supplementary Discussion), the estimated ¹⁴C ages ranged from about 2,900 to 12,100 years (Table 1).

The geochemical, isotopic and molecular data indicate that the nematode-bearing water represents palaeometeoric, hypoxic water that contains a microbial assemblage comprising both aerobic and anaerobic bacteria. The cultured nematodes preferentially fed on these bacteria as opposed to *E. coli*. Nematodes were absent from the mining water, and when found twice in soils they were taxonomically distinct from those found in the palaeometeoric water. No other metazoans were found in the palaeometeoric water. The nematodes, therefore, do not seem to be contaminants due to mining or incursion of modern water, but seem to be indigenous to the palaeometeoric water.

The subsurface microbial population density can sustain nematodes for thousands of years. The ratio of planktonic microbial cells to nematodes was 10⁸–10¹⁰:1 (Table 1), greatly exceeding the 10–100:1 ratio of microbial cells to protists reported for other terrestrial subsurface environments⁶. Rather than feeding on planktonic cells, the nematodes

are more likely to graze on the patches of bacteria (~5 × 10⁴ cells per square centimetre) attached to the fracture surfaces¹⁷. This density corresponds to a sessile microbial concentration 100 times that of the planktonic cells, making the ratio of microbial cells to nematodes ~10¹⁰–10¹²:1. Given the dry weight of *H. mephisto*, 2.6 × 10⁻⁸ g (Supplementary Discussion), measured bacterivory rates of 6.6 × 10⁵ to 15.2 × 10⁵ bacterial cells per millimetre of nematode per day and respiration/total carbon consumption ratios of 0.16 to 0.72 (ref. 18) for nematodes, ~10⁴ bacterial cells could readily sustain *H. mephisto* for one day and 10¹¹ for about 30,000 years. The association of nematodes with biofilms may explain why they were detected only when the boreholes were first opened and not in subsequent samples. The initial release of high-pressure water would have dislodged biofilms and nematodes from the fracture surfaces and some time would be required for the biofilm community to reform after the borehole was sealed again. This lack of reproducibility in eukaryote presence was also observed for subsurface fungi⁶.

Dioxygen is not a limiting factor. Water samples yielded O₂ concentrations of 13 to 72 μM (equivalent to 1.3 to 6.8 kPa of O₂ in Table 1), values that can sustain the maximum metabolic rate of 3 × 10⁻⁴ moles of O₂ per gram of nematode per hour for *Caenorhabditis elegans*^{19,20}. If the palaeometeoric water had an O₂ concentration of ~350 μM on recharge, then the ¹⁴C ages constrain the O₂ consumption rate to be ~3 × 10⁻⁸ to 8 × 10⁻⁸ M yr⁻¹, values which overlap those reported for the Middelndorf aquifer²¹. Given the mass of *H. mephisto*, its maximum metabolic rate would be ~7 × 10⁻⁸ moles of O₂ per nematode per year. The observed O₂ consumption rate could support ~1 nematode per litre at its maximum metabolic rate, compared to the 1 nematode per 10⁴ litres detected. Even at an O₂ concentration of 4 μM, nematodes metabolize aerobically⁹, but at much slower rates¹⁹.

Temperature imposes a limit to the depth at which nematodes could live. The temperature of Beatrix borehole 2 is higher than most terrestrial nematodes can tolerate, but *H. mephisto* showed a high temperature tolerance (Table 2) like that of the opportunistic pathogen *Halicephalobus gingivalis*. The detection of the monhysterid species in the Tau Tona borehole at 48 °C is significantly higher than a reported occurrence of this species in a hot spring at 43 °C (ref. 22), but is

Table 2 | Nematode characteristics

Species	Maximum growth temperature (°C)	Reproduction	Feeding type
<i>H. mephisto</i>	41	Parthenogenetic	Bacteriophagous
<i>P. aquatilis</i>	31	Parthenogenetic	Bacteriophagous
Monhysterid sp. 1	Unknown	Sexual?	Bacteriophagous
Monhysterid sp. 2	Unknown	Unknown	Bacteriophagous*

* Monhysteridae are classified as 'bacterial feeding' and/or 'substrate ingestion' on the basis of their buccal morphology²⁹. Substrate ingestion is theoretically identical to bacterial feeding, predation and unicellular eukaryote feeding in many groups, because more than a pure food source is ingested.

comparable to the occurrences of nematodes in hot springs at 51 to 61 °C (refs 23–25).

Our data suggest that the interactions between meiofaunal communities and chemolithotrophic biofilms found in hypoxic, sulphide cave systems, for example Movile cave in Romania²⁶, extend to greater depths and smaller confines and could control the size and turnover rate of subsurface microbial communities. Given their abundance on the sea floor and around hydrothermal vents, nematodes should also be found beneath the sea floor in the sediments or in mid-ocean-ridge basalt²⁷. Other meiofauna, such as the Loricifera inhabiting the deep, anoxic, hypersaline L'Atalante basin²⁸, could also be present. The ability of multicellular organisms to survive in the subsurface should be considered in the evolution of eukaryotes and the search for life on Mars.

Received 15 February; accepted 1 March 2011.

- Pedersen, K. The deep subterranean biosphere. *Earth Sci. Rev.* **34**, 243–260 (1993).
- Onstott, T. C. *et al.* in *Enigmatic Microorganisms and Life in Extreme Environments* (ed. Seckbach, J.) 487–500 (Kluwer, 1998).
- Amend, J. P. & Teske, A. Expanding frontiers in deep subsurface microbiology. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **219**, 131–155 (2005).
- Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci. USA* **95**, 6578–6583 (1998).
- Sinclair, J. L. & Ghiorse, W. C. Distribution of aerobic bacteria, protozoa, algae and fungi in deep subsurface sediments. *Geomicrobiol. J.* **7**, 15–31 (1989).
- Ekendahl, S., O'Neill, A., Thomsson, E. & Pedersen, K. Characterisation of yeasts isolated from deep igneous rock aquifers of the Fennoscandian shield. *Microb. Ecol.* **46**, 416–428 (2003).
- Heip, C., Vincx, M. & Vranken, G. The ecology of marine nematodes. *Oceanogr. Mar. Biol.* **23**, 399–489 (1985).
- Lambshead, P. in *Nematode Morphology, Physiology and Ecology* Vol. 1 (eds Chen, Z. X., Chen, S. Y. & Dickson, D. W.) 438–492 (Tsinghua Univ. Press, 2004).
- Föll, R. L. *et al.* Anaerobiosis in the nematode *Caenorhabditis elegans*. *Comp. Biochem. Physiol.* **124B**, 269–280 (1999).
- Moser, D. P. *et al.* *Desulfotomaculum* spp. and *Methanobacterium* spp. dominate 4–5 km deep fault. *Appl. Environ. Microbiol.* **71**, 8773–8783 (2005).
- Dorris, M., De Ley, P. & Blaxter, M. Molecular analysis of nematode diversity and the evolution of parasitism. *Parasitol. Today* **15**, 188–193 (1999).
- Holterman, M. *et al.* Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown clades. *Mol. Biol. Evol.* **23**, 1792–1800 (2006).
- Lippmann, J. *et al.* Dating ultra-deep mine waters with noble gases and ³⁶Cl, Witwatersrand Basin, South Africa. *Geochim. Cosmochim. Acta* **67**, 4597–4619 (2003).
- Onstott, T. C. *et al.* The origin and age of biogeochemical trends in deep fracture water of the Witwatersrand Basin, South Africa. *Geomicrobiol. J.* **23**, 369–414 (2006).
- Gihring, T. M. *et al.* The distribution of microbial taxa in the subsurface water of the Kalahari Shield, South Africa. *Geomicrobiol. J.* **23**, 415–430 (2006).
- Michel, R. L. in *Isotopes in the Water Cycle: Past, Present and Future of a Developing Science* (eds Aggarwal, P. K., Gat, J. R. & Froelich, K. F. O.) Ch. 5, 53–66 (Springer, 2005).
- Wanger, G., Onstott, T. C. & Southam, G. Structural and chemical characterization of a natural fracture surface from 2.8 kilometers below land surface: biofilms in the deep subsurface. *Geomicrobiol. J.* **23**, 443–452 (2006).
- Ferris, H., Venette, R. C. & Lau, S. S. Population energetics of bacterial-feeding nematodes: carbon and nitrogen budgets. *Soil Biol. Biochem.* **29**, 1183–1194 (1997).
- van Voorhies, W. & Ward, S. Broad oxygen tolerance in the nematode *Caenorhabditis elegans*. *J. Exp. Biol.* **203**, 2467–2478 (2000).
- Van Voorhies, W. A. & Ward, S. Genetic and environmental conditions that increase longevity in *Caenorhabditis elegans* decrease metabolic rate. *Proc. Natl Acad. Sci. USA* **96**, 11399–11403 (1999).
- Phelps, T. J., Murphy, E. M., Pfiffner, S. M. & White, D. C. Comparison between geochemical and biological estimates of subsurface microbial activities. *Microb. Ecol.* **28**, 335–349 (1994).
- Ocana, A. Relationship between nematode species and the physico-chemical characteristics of spring waters. II. Temperature. *Nematol. Mediterr.* **19**, 25–28 (1991).
- Neher, D. A. & Powers, T. O. in *Encyclopedia of Soils in the Environment* Vol. 3 (eds Hillel, D. *et al.*) 1–5 (Academic, 2004).
- Hoeppli, R. & Chu, H. J. Free-living nematodes from hot springs in China and Formosa. *Hong Kong Nat.* **1** (suppl.), 15–29 (1932).
- Jana, B. B. The thermal springs of Bakreswar, India: physico-chemical conditions, flora and fauna. *Hydrobiologia* **41**, 291–307 (1973).
- Engel, A. S. Observations on the biodiversity of sulfidic karst habitats. *J. Cave Karst Stud.* **69**, 187–206 (2007).
- Edwards, K. J., Bach, W. & McCollom, T. M. Geomicrobiology in oceanography: microbe–mineral interactions at and below the seafloor. *Trends Microbiol.* **13**, 449–456 (2005).
- Danovaro, R. *et al.* The first metazoa living in permanently anoxic conditions. *BMC Biol.* **8**, 30–40 (2010).
- Yeates, G. W., Bongers, T., Goede, R. G. M., d., Freckman, D. W. & Georgieva, S. S. Feeding habits in nematode families and genera — an outline for soil ecologists. *J. Nematol.* **25**, 315–331 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements G.B. is grateful to the University of Ghent for allowing a year-long sabbatical, and for grants from the FWO, BOF and UFS. We express our recognition of the support provided by Tau Tona gold mine (AngloGold Ashanti Limited), Driefontein gold mine (Gold Fields Limited), Beatrix gold mine (Gold Fields Limited), Zondereinde mine (Northern Platinum Ltd) and Star Diamonds mine (Petra Diamonds), and by the many people without whom this work would have been impossible. In particular, we are grateful to F. Rheeder, H. Möller, T. Lineque, A. Thwala, K. Sokhela, C. Rose, R. Fynn, B. Visser, O. Holovachev, T. Moens, M. Couvreur and A. Vierstraete. We are grateful to E. Botes and K. Albertyn of the University of the Free State for their contributions to the data analysis. T.C.O. acknowledges support from a National Science Foundation Continental Dynamics Program grant (EAR 0409605). E.v.H. acknowledges support from a BioPAD/UFS Metagenomics Platform grant.

Author Contributions A.G.-M., D.L. and W.B. all contributed equally to this study. G.B., A.G.-M., D.L., A.B. and M.E. collected the filtered samples and the control samples and performed field analyses. G.B. carried out the enrichments. A.G.-M. performed microbial DNA extraction and 16S rRNA amplification, sequencing and tree construction. C.M. performed DNA analyses on filters of mining water. W.B. provided the nematode identification, their morphological description and their molecular analyses. T.C.O. modelled the geochemical, ³H and ¹⁴C data. G.B. wrote the paper with input from W.B., A.G.-M., T.C.O. and E.v.H.

Author Information Sequence information for *H. mephisto* has been deposited at GenBank under accession number GQ918144. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.B. (gborgonie@gmail.com) or T.C.O. (tullis@princeton.edu).

Earliest evidence of mammalian social behaviour in the basal Tertiary of Bolivia

Sandrine Ladevèze¹, Christian de Muizon², Robin M. D. Beck³, Damien Germain² & Ricardo Cespedes-Paz⁴

The vast majority of Mesozoic and early Cenozoic metatherian mammals (extinct relatives of modern marsupials) are known only from partial jaws or isolated teeth, which give insight into their probable diets and phylogenetic relationships but little else. The few skulls known are generally crushed, incomplete or both^{1–4}, and associated postcranial material is extremely rare. Here we report the discovery of an exceptionally large number of almost undistorted, nearly complete skulls and skeletons of a stem-metatherian, *Pucadelphys andinus*, in the early Palaeocene epoch⁵ of Tiupampa in Bolivia^{6–8}. These give an unprecedented glimpse into early metatherian morphology, evolutionary relationships and, especially, ecology. The remains of 35 individuals have been collected, with 22 of these represented by nearly complete skulls and associated postcrania. These individuals were probably buried in a single catastrophic event, and so almost certainly belong to the same population⁹. The preservation of multiple adult, sub-adult and juvenile individuals in close proximity ($<1\text{ m}^2$) is indicative of gregarious social behaviour or at least a high degree of social tolerance and frequent interaction. Such behaviour is unknown in living didelphids, which are highly solitary and have been regarded, perhaps wrongly, as the most generalized living marsupials. The Tiupampa *P. andinus* population also exhibits strong sexual dimorphism, which, in combination with gregariousness, suggests strong male–male competition and polygyny. Our study shows that social interactions occurred in metatherians as early as the basal Palaeocene and that solitary behaviour may not be plesiomorphic for Metatheria as a whole.

The locality of Tiupampa, Bolivia^{10,11}, has yielded an abundant and remarkably preserved mammal fauna from the earliest Palaeocene⁶. Some of the most spectacular fossils are numerous partial skulls and skeletons of the small (about 20 g) stem-metatherian, *P. andinus* (Fig. 1 and Supplementary Information). In 1985, four partial skeletons (two of them with a skull) and eight additional sub-complete to partial skulls were discovered in an area of less than a square metre. Some isolated jaws and teeth were also collected in this area. In 1997, a second concentration of *P. andinus* remains was discovered approximately 3 m away from the first area, with 23 skulls and associated postcranial skeletons present in approximately a square metre (Fig. 1). These previously undescribed 1997 specimens have now been prepared and, together with the 1985 specimens, provide significant new information about the ecology of *Pucadelphys*.

In both areas, the skeletons were often intermixed with little or no disarticulation. Such preservation demonstrates that there was no postmortem transportation. The abundance of sub-complete articulated skeletons of other small vertebrates (pantodonts, two other metatherians, frogs, lizards, snakes, newborn crocodiles) and the presence of numerous crocodile eggs (including an intact nest) all within a very restricted area (less than 100 m^2) suggests that a single catastrophic event buried this fauna *in situ*; sedimentological evidence indicates that the fauna existed by a pond or ox-bow lake, adjacent

to the bank of a large meandering river^{9,10,12}. *P. andinus* appears to have been mainly terrestrial with arboreal abilities¹³, and probably lived in burrows¹⁰. In this context, it is highly probable that the *P. andinus* individuals were living together and therefore comprise a sample of the same population. Of the 35 specimens represented by partial skulls and/or partial skeletons, 22 present exceptionally preserved skulls and mandibles (12 adult females, six adult males, four sub-adults), which is unprecedented in the fossil record of pre-Neogene therians. Indeed, *Pucadelphys* is now known from more specimens than several living marsupials (for example, *Glironia venusta*, *Microperoryctes murina*).

Four of the skulls of the *P. andinus* sample represent sub-adults, given that their adult dentition is not fully erupted: P3 (third upper premolar) and m4 (fourth lower molar) are both partially erupted, while tooth M4 is unerupted. From comparison with extant didelphid marsupials, this dental stage may correspond to either immature or sexually precocious individuals^{14,15}. A single lower jaw¹⁶ retains an alveolus for tooth dp3 (third deciduous lower premolar), and tooth m3 is only partially erupted, suggesting that it is from an immature individual^{14,15}. Of the 18 skulls with fully erupted adult dentition, six are considerably larger than the others and exhibit longer canines and higher sagittal and occipital crests. However, the size, relative proportions and morphology of the postcanine dentition indicate that all 18 individuals belong to a single species. The morphological differences are here interpreted as evidence of strong sexual dimorphism (Fig. 2).

Body-mass estimates (Supplementary Information) indicate that the larger skulls, considered here to represent males, are from individuals that weighed on average 35% more than the smaller-skulled (female) individuals (Supplementary Information). Strong sexual size dimorphism is found in some extant marsupials that are probably ecologically similar to *P. andinus* in terms of body size, diet (based on dental morphology) and pattern of locomotion (terrestrial–scansorial to arboreal), namely small South American didelphids (for example, species of *Gracilinanus*, *Marmosa*, *Marmosops* and *Monodelphis*¹⁷) and small Australian dasyurids (for example, *Antechinus*¹⁸). However, the dimorphism seen in these living taxa appears to be connected with semelparity (all males die shortly after a single breeding season)^{17,19}, and *P. andinus* was clearly not semelparous, given the coexistence of adult males with sub-adults and juveniles. Sexual size dimorphism in small dasyurids may be related to male dominance over females, which are generally territorial and commonly aggressive towards conspecifics²⁰. However, the presence of multiple adult males and females in the *P. andinus* sample indicates that females were non-territorial, and so the observed sexual size dimorphism is more likely to be connected with male–male competition, and hence polygyny (one male mates with more than one female).

Principal component analyses (PCAs) based on a maximum of 43 craniodental measurements reveal good discrimination of putative male and female individuals, as well as between adults and sub-adults (Fig. 3, Supplementary Information). A sexual dimorphism index (SDI,

¹Department of Paleontology, Royal Belgian Institute of Natural Sciences, 29 rue Vautier, B-1000 Brussels, Belgium. ²Département Histoire de la Terre UMR7207 CR2P (CNRS, MNHN, UPMC), 8 rue Buffon, F-75005 Paris, France. ³Department of Mammalogy, American Museum of Natural History, Central Park West 79th Street, New York, New York 10024, USA. ⁴Museo de Historia Natural Alcide D'Orbigny, Av. Potosi N-1458, Cochabamba, Bolivia.

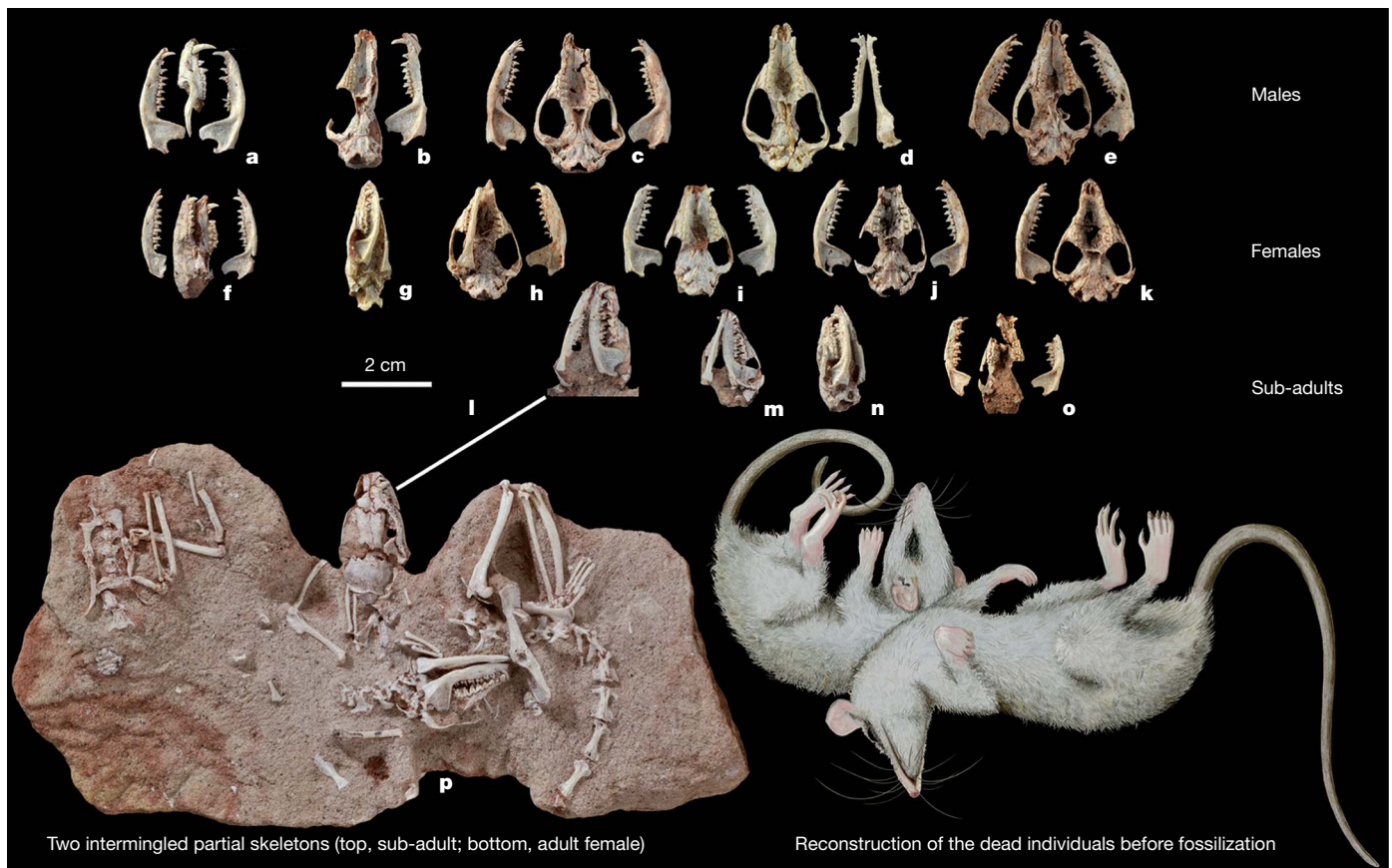


Figure 1 | Skulls and partial skeletons of *P. andinus* from the same population of basal Palaeocene beds at Tiupampa (Bolivia). a, MHNC 8364. b, MHNC 8377. c, MHNC 8381. d, MHNC 8382. e, MHNC 8266. f, MHNC 8386. g, MHNC 8389. h, MHNC 8379. i, MHNC 8376. j, MHNC 8380. k, MHNC 8378. l, MHNC 8391. m, MHNC 8388. n, MHNC 8265. o, MHNC

8384. p, MHNC 8392. The differences of position in skulls and bones between the block with tangled skeletons (left) and the reconstruction (right) is interpreted as bone displacement after flesh decomposition and prior to sediment consolidation.



as defined in the Supplementary Information), together with identification of the most dimorphic parameters by analyses of variance (ANOVAs), reveals that *P. andinus* males are characterized by longer and wider skulls (markedly expanded posteriorly), longer and higher dentaries and larger canines. These sexually dimorphic craniodental features have been shown to be reliable indicators of social structure in living faunivorous mammals²¹, and further supports the existence of strong male–male competition and polygyny in *P. andinus*.

Beyond these sexually dimorphic morphological features, the *P. andinus* sample also reveals extreme intra-specific and intra-individual variability in molar morphology that is unrelated to sexual dimorphism. Indeed, if found as isolated elements, several teeth would probably have been identified as representing different taxa. The most variable molar features are overall robustness and size, morphology of the styler cusps, protocone width, talonid labial angle, entoconid size, paraconid position and, most strikingly, the overall morphology and proportion of the ultimate molars (Fig. 4). Such intra-population variation in molar morphology has not been observed in similar-sized modern didelphids (except for the styler cusp variation²²), but is present in some populations of the much larger *Didelphis virginiana*.

The population structure seen in the *Pucadelphys* fossil assemblage comprises multiple adults of both sexes and different ages (based on the degree of tooth wear), together with sub-adults and one juvenile. Extant didelphids are solitary animals occurring at low overall population densities and are usually highly intolerant of conspecifics²³.

Figure 2 | Skulls and mandibles of male (right, MHNC 8266) and female (left, MHNC 8378) of *P. andinus*. Top, dorsal view; middle, ventral view; bottom, lateral view.

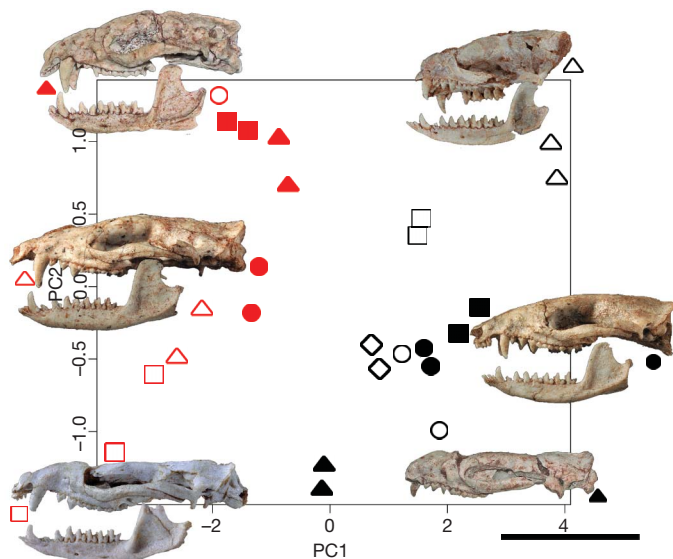


Figure 3 | PCA of 32 cranial and dental variables in 13 adult specimens of *P. andinus*. (This is PCA3; see Supplementary Information for the method and other PCAs). The repartition of the variables on the two first axes of the PCA highlights a good discrimination of the males (left, in red) and females (right, in black). The symbols refer to specimens: open black circle, MHNC 8376; open red circle, MHNC 8377; filled black circle, MHNC 8378; open black square, MHNC 8379; filled black square, MHNC 8380; filled red circle, MHNC 8381; open red square, MHNC 8382; filled red square, MHNC 8364; open red triangle, MHNC 8266; open black triangle, MHNC 8386; filled red triangle, YPFB Pal 6105; filled black triangle, YPFB Pal 6108; open black diamond, YPFB Pal 6110. Scale, 2 cm.

Female didelphids are highly territorial and their home ranges rarely overlap. Most dasyurids are also solitary²⁴. By contrast, the presence of the fossils of at least 35 individuals, including 12 adult females, within an area of a few square metres strongly suggests that *P. andinus* was gregarious and that females were non-territorial.

Available data suggest that didelphids have a non-monogamous, probably promiscuous, mating system, with males actively searching

for females during the breeding season²⁵. Even during the breeding season, no didelphid species has been reported to form groups as large as those inferred in *P. andinus*. Didelphids never live in groups that include adults of both sexes together with sub-adults and juveniles, again unlike *P. andinus*. Dasyurid mating systems are either based on promiscuous matings between solitary individuals in overlapping home ranges²³ or lek promiscuity, with individual females visiting aggregations of males²⁶. The presence of multiple sub-adults and at least one juvenile in the *P. andinus* sample, together with the female-biased sex ratio of 12 adult females and six adult males, suggests that the inferred gregariousness in this species is not connected to mating.

Although most dasyurids are predominantly or entirely solitary, nesting groups occur in some species. These usually comprise a mother with young and postweaning juveniles, but nesting groups of between two and 40 individuals have been observed²⁶. However, such large nesting groups have only been reported in southern Australia, and communal nesting only occurs during winter with the onset of very cold nights²⁶. This pattern, together with the observation that the dasyurid *Sminthopsis crassicaudata* also forms winter nests with the house mouse *Mus musculus* (normally its prey)²⁷, suggests that the primary purpose of large nesting groups is thermoregulatory, with huddling allowing energetic savings. Bolivia in the early Cenozoic was approximately at the same latitude as it is now (the present latitude of Tiupampa is about 18° S) but at much lower altitude than today. The climate at Tiupampa was probably similar to that of the modern Amazonian rainforest, namely tropical with temperatures well above freezing throughout the year and with wet–dry seasonality, which is reinforced by the fact that the neotropical temperatures during the Palaeocene were significantly warmer than they are now (about +6 °C warmer)²⁸. The gregariousness inferred in *P. andinus* was therefore unlikely to be connected with thermoregulation.

In summary, the fossils discussed here demonstrate that the early Palaeocene *P. andinus* population at Tiupampa was a socially interacting group, characterized by gregariousness, male–male competition and polygyny. The probable tropical climate of Tiupampa during the early Palaeocene suggests that breeding in *P. andinus* was probably aseasonal, which may have encouraged male–male competition and hence led to the evolution of sexual size and shape dimorphism. The discovery at Tiupampa of six articulated and intermingled skeletons of another metatherian taxon, *Andinodelphys cochabambensis* (which was probably more arboreal than *Pucadelphys*)¹³, may indicate gregariousness in this taxon as well. The Tiupampa metatherians therefore indicate for the first time that, as early as the basal Palaeocene, gregarious social behaviour was present in metatherians, and that the highly solitary behaviour of living didelphids may be a derived rather than primitive characteristic.

METHODS SUMMARY

The detailed method and calculations for estimating the body mass of *P. andinus* is provided in the Supplementary Information. Multivariate analyses were performed (PCAs and ANOVAs) and are described in the Supplementary Information. The quantification of the sexual dimorphism was possible thanks to a new index, the description of which is given in the Supplementary Information. Also see Supplementary Information for detailed Methods, data sets, measurements, calculations and further references.

Received 17 January; accepted 3 March 2011.

Published online 8 May; corrected 2 June 2011 (see full-text HTML version for details).

1. Luo, Z.-X., Ji, Q., Wible, J. R. & Yuan, C.-X. An early Cretaceous tribosphenic mammal and metatherian evolution. *Science* **302**, 1934–1940 (2003).
2. Rougier, G. W., Wible, J. R. & Novacek, M. J. Implications of *Deltatheridium* specimens for early marsupial history. *Nature* **396**, 459–463 (1998).
3. Szalay, F. S. & Trofimov, B. The Mongolian Late Cretaceous *Asiatherium*, and the early phylogeny and paleobiogeography of Metatheria. *J. Vert. Pal.* **16**, 474–509 (1996).
4. Horowitz, I. et al. Cranial anatomy of the earliest marsupials and the origin of opossums. *PLoS ONE* **4**, e8278 (2009).

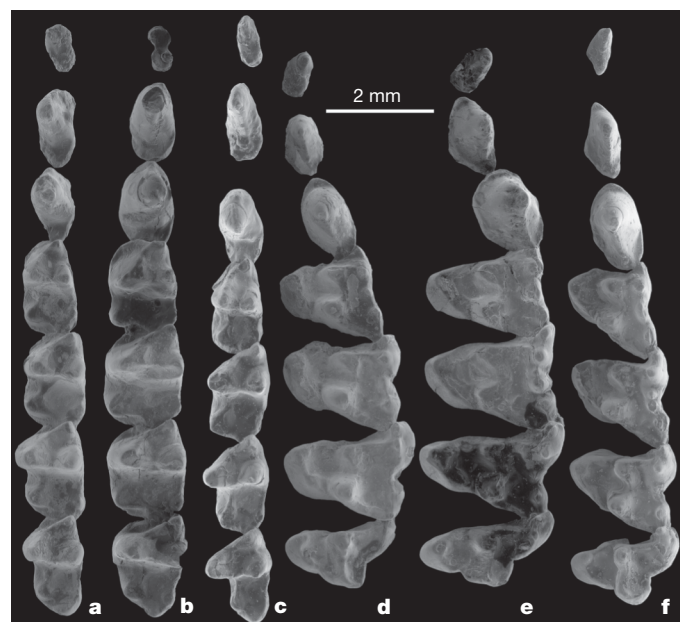


Figure 4 | Upper and lower dental series of *P. andinus* to show the individual variation of dental morphology. **a, d**, Left lower (**a**) and left upper (**d**) tooth row of MHNC 8378 (female). **b, e**, Right (reversed) lower (**b**) and right (reversed) upper (**e**) tooth row of MHNC 8381 (male). **c, f**, Right (reversed) lower (**c**) and right (reversed) upper (**f**) tooth row of MHNC 8266 (male).

5. Gelfo, J. N., Goin, F. J., Woodburne, M. O. & de Muizon, C. Biochronological relationships of the earliest South American Paleogene mammalian faunas. *Palaeontology* **52**, 251–269 (2009).
6. Marshall, L. G. & de Muizon, C. in *Pucadelphys andinus* (Marsupialia, Mammalia) from the early Paleocene of Bolivia. Phylogenetic and palaeobiologic implications. *Muséum national d'Histoire naturelle*, 1995).
7. de Muizon, C. *Mayulestes ferox*, a boryaenoid (Metatheria, Mammalia) from the early Palaeocene of Bolivia. *Nature* **389**, 486–489 (1997).
8. de Muizon, C., Cifelli, R. L. & Céspedes Paz, R. The origin of the dog-like borhyaenoid marsupials of South America. *Nature* **389**, 486–489 (1997).
9. de Muizon, C. & Céspedes, R. in *Origins and Evolution of Cenozoic South American Mammals* (eds Rosenberger, A. & Tejedor, M.) (Springer, in the press).
10. Marshall, L. G., de Muizon, C. & Sigogneau-Russel, D. *Pucadelphys andinus* (Marsupialia, Mammalia) from the Early Paleocene of Bolivia (Mémoires du Muséum national d'Histoire naturelle, 1995).
11. de Muizon, C. in *Fósiles y Fácies de Bolivia* Vol. I *Vertebrados* Vol. 12 (3–4) (ed. Suárez-Soruco, R.) 575–624 (Revista Técnica, Yacimientos Petrolíferos Fiscales Bolivianos, 1992).
12. Marshall, L. G., Sempere, T. & Butler, R. F. Chronostratigraphy of the mammal-bearing Paleocene of South America. *J. S. Am. Earth Sci.* **10**, 49–70 (1997).
13. de Muizon, C. & Argot, C. in *Predators with Pouches: the Biology of Carnivorous Marsupials* (eds Jones, M., Dickman, C. & Archer, M.) 42–63 (Surrey Beatty & Sons, 2003).
14. Atramementowicz, M. Dynamique de population chez trois marsupiaux didelphidés de Guyane. *Biotropica* **18**, 136–149 (1986).
15. Tyndale-Biscoe, C. H. & MacKenzie, R. B. Reproduction in *Didelphis marsupialis* and *Didelphis albiventris* in Colombia. *J. Mamm.* **57**, 249–265 (1976).
16. Cifelli, R. L. & de Muizon, C. Tooth eruption and replacement pattern in early marsupials. *C. R. Acad. Sci. Ser. D* **326**, 215–220 (1998).
17. Astúa, D. Cranial sexual dimorphism in New World marsupials and a test of Rensch's rule in Didelphidae. *J. Mamm.* **91**, 1011–1024 (2010).
18. Fisher, D. O. & Cockburn, A. The large-male advantage in brown antechinuses: female choice, male dominance, and delayed male death. *Behav. Ecol.* **17**, 164–171 (2005).
19. Kraaijeveld, K., Kraaijeveld-Smit, F. J. L. & Adcock, G. J. Does female mortality drive male semelparity in dasyurid marsupials? *Proc. R. Soc. Lond. B* **270** (Suppl. 2), S251–S253 (2003).
20. Soderquist, T. Spatial organization of the arboreal carnivorous marsupial, *Phascogale tapoatafa*. *J. Zool.* **237**, 385–398 (1995).
21. Plavcan, J. M. Inferring social behavior from sexual dimorphism in the fossil record. *J. Hum. Evol.* **39**, 327–344 (2000).
22. Hershkovitz, P. Composition of the family Didelphidae Gray, 1821 (Didelphoidea: Marsupialia), with a review of the morphology and behavior of the included four-eyed pouched opossums of the genus *Philander* Tiedemann, 1808. *Feldiana Zool.* **86**, 1–103 (1997).
23. Russell, E. M. Social behaviour and social organization of marsupials. *Mammal Rev.* **14**, 101–154 (1984).
24. Oakwood, M. Spatial and social organization of a carnivorous marsupial *Dasyurus hallucatus* (Marsupialia: Dasyuridae). *J. Zool.* **257**, 237–248 (2002).
25. Cáceres, N. C., Napoli, R. P., Lopes, W. H., Casella, J. & Gazeta, G. S. Natural history of the marsupial *Thylamys macrurus* (Mammalia, Didelphidae) in fragments of savannah in southwestern Brazil. *J. Nat. Hist.* **41**, 1979–1988 (2007).
26. Lazenby-Cohen, K. A. & Cockburn, A. Lek promiscuity in a semelparous mammal, *Antechinus stuartii* (Marsupialia: Dasyuridae)? *Behav. Ecol. Sociobiol.* **22**, 195–202 (1988).
27. Morton, S. R. Torpor and nest-sharing in free-living *Sminthopsis crassicaudata* (Marsupialia) and *Mus musculus* (Rodentia). *J. Mamm.* **59**, 569–575 (1978).
28. Head, J. J. *et al.* Giant boid snake from the Palaeocene neotropics reveals hotter past equatorial temperatures. *Nature* **457**, 715–717 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Field expeditions were supported by the National Geographic Society (grants 2908-84) in 1985 and by the 'Institut Français d'Études Andines' in 1997. The specimens collected are the property of the 'Museo de Historia Natural Alcide d'Orbigny' in Cochabamba, Bolivia. Additional financial support was provided by Research Project MO/36/020 of the Belgian Federal Science Policy Office (S.L.), and NSF grant DEB-0743039 (R.M.D.B.), in collaboration with R. Voss at the American Museum of Natural History. Photographs are by C. Lemzaouda and P. Loubry, scanning electron microscope photographs are by J. Cillis and the reconstruction in Fig. 1 is by S. Fernandez. We thank E. Westwig for access to the AMNH collections, and O. Lambert, A. Pradel, and T. Soderquist for discussions.

Author Contributions C.d.M. and R.C.-P. collected the specimens at Tiupampa. C.d.M. prepared the specimens. S.L. and C.d.M. initiated and organized the project. S.L., C.d.M. and R.M.D.B. wrote the paper. S.L., C.d.M., R.M.D.B., D.G. and R.C.-P. discussed the results and commented on the manuscript at all stages. Measurements were taken by S.L. and R.M.D.B. PCAs and ANOVAs were performed by S.L. and D.G.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.L. (sandrine.ladeveze@naturalsciences.be) or C.M. (muizon@mnhn.fr).

Ultra-prolonged activation of CO₂-sensing neurons disorients mosquitoes

Stephanie Lynn Turner^{1*}, Nan Li^{2*}, Tom Guda³, John Githure³, Ring T. Cardé² & Anandasankar Ray^{1,2}

Carbon dioxide (CO₂) present in exhaled air is the most important sensory cue for female blood-feeding mosquitoes, causing activation of long-distance host-seeking flight, navigation towards the vertebrate host¹ and, in the case of *Aedes aegypti*, increased sensitivity to skin odours². The CO₂ detection machinery is therefore an ideal target to disrupt host seeking. Here we use electrophysiological assays to identify a volatile odorant that causes an unusual, ultra-prolonged activation of CO₂-detecting neurons in three major disease-transmitting mosquitoes: *Anopheles gambiae*, *Culex quinquefasciatus* and *A. aegypti*. Importantly, ultra-prolonged activation of these neurons severely compromises their ability subsequently to detect CO₂ for several minutes. We also identify odours that strongly inhibit CO₂-sensitive neurons as candidates for use in disruption of host-seeking behaviour, as well as an odour that evokes CO₂-like activity and thus has potential use as a lure in trapping devices. Analysis of responses to panels of structurally related odours across the three mosquitoes and *Drosophila*, which have related CO₂-receptor proteins, reveals a pattern of inhibition that is often conserved. We use video tracking in wind-tunnel experiments to demonstrate that the novel ultra-prolonged activators can completely disrupt CO₂-mediated activation as well as source-finding behaviour in *Aedes* mosquitoes, even after the odour is no longer present. Lastly, semi-field studies demonstrate that use of ultra-prolonged activators disrupts CO₂-mediated hut entry behaviour of *Culex* mosquitoes. The three classes of CO₂-response-modifying odours offer powerful instruments for developing new generations of insect repellents and lures, which even in small quantities can interfere with the ability of mosquitoes to seek humans.

Each year, *A. gambiae*, *A. aegypti* and *C. quinquefasciatus* mosquitoes transmit deadly diseases such as malaria and dengue to more than half a billion people and cause millions of deaths. Female mosquitoes need to find and obtain a blood meal from an infected person and subsequently find an uninfected human host to bite. Disruption of host-seeking in mosquitoes could therefore reduce incidence of both steps.

Female mosquitoes depend primarily on olfactory cues that are emitted from human breath, skin and sweat to find hosts³. Yet very few strategies that target the mosquito olfactory system, like the effective repellent *N,N*-diethyl-*m*-toluamide (DEET)^{4,5}, have emerged. However, the relatively high cost of DEET and the need for repeated application to the skin at high concentrations (10–70%) precludes its use in tropical countries. DEET is also known to melt plastics⁶, and to block cation channels⁴ and cholinesterase activity in mammals⁷. Moreover, DEET resistance has been observed in *Drosophila melanogaster*⁸, *Anopheles albimanus*⁹ and *A. aegypti*¹⁰, which reiterates a need to identify additional classes of repellents.

An ideal target for disrupting host seeking is the CO₂-detection machinery. CO₂ exhaled by humans is a vital cue for mosquitoes. Female mosquitoes are exquisitely sensitive to minute changes in CO₂ concentration¹¹, which they potentially sense from long distances^{1,12}. Upon encountering a plume of CO₂, a female mosquito orients upwind

using optomotor anemotaxis, while maintaining contact with the host-odour plume¹³. Remarkably, CO₂ exposure also increases the sensitivity of *A. aegypti* to human skin odours².

CO₂ is detected by heteromeric receptor proteins that are highly conserved across dipterans including the three vector mosquitoes studied here and *Drosophila*¹⁴ (Supplementary Fig. 1a), and are expressed in a dedicated labelled-line CO₂-detection circuit^{15–18}. This circuit provides the means to disrupt key aspects of host-seeking behaviour in multiple species. Previously we identified odorants that inhibit electrophysiological and behavioural responses to CO₂ in *D. melanogaster*, two of which also inhibited the CO₂ response in *C. quinquefasciatus* mosquitoes¹⁹.

To identify more effective and across-species inhibitors of conserved CO₂ receptors¹⁴, we screened a panel of compounds structurally related to known inhibitors using single-sensillum electrophysiology on *A. gambiae*, *A. aegypti* and *C. quinquefasciatus* females, and identified several efficient inhibitors for each (Fig. 1a). The effects of some compounds were conserved across 250 million years of evolution²⁰ whereas others were species specific. Subtle changes in structure influenced the strength of inhibition (Fig. 1a), but volatility or solubility did not seem to be as important (Supplementary Fig. 2). Inhibition was dose dependent (Fig. 1b, c) and the temporal dynamics of inhibition were odour dependent (Supplementary Fig. 1b).

Most trapping devices use CO₂ to attract mosquitoes. Unfortunately, these are rarely used for surveillance or control in tropical countries because of high costs and difficulties in supplying CO₂ via gas cylinders, dry ice, or propane combustion. Using a parallel electrophysiology screen we identified 2-butanone as a dose-dependent activator of the cpA neuron in all three mosquito species (Supplementary Fig. 3a, b). Repeated stimuli of 2-butanone showed a temporal pattern of cpA neuron activation indistinguishable from that elicited by CO₂ (Fig. 1d, e). Such odours, which mimic CO₂ activity, are ideal candidates for formulating convenient, compact and economical lures for use in trap-based mosquito control in developing countries.

In the electrophysiology screen we also identified an additional structurally related odorant, 2,3-butanedione, which activates the cpA neuron in the three mosquito species (Supplementary Fig. 4a). 2,3-Butanedione exhibits an unusual property: exposure to a brief 1-s pulse results in ultra-prolonged activation of the cpA neuron (Fig. 1f–h). The activity of the cpA neuron remained high for the entire 5.5-min recording period. In contrast, response to a comparable CO₂ stimulus rapidly decays immediately after the end of stimulus (Fig. 1g, h). Olfactory receptor neurons (ORNs) that express *Or* genes rarely show this property and rapidly adapt in the first second upon stimulus exposure, even with a continuous stimulus²¹. Not much has been reported quantitatively on temporal response termination kinetics for single odorants beyond two seconds post stimulus²². An individual neuron trace of a moth to a plant volatile suggests activation beyond the stimulus period²³. In another example, an ORN of a male moth showed activity for >10 min after a 1-s pulse of a pheromone analogue, but only with an increase of ~3 spikes s⁻¹ (ref. 24).

¹Cellular, Molecular, and Developmental Biology Program, University of California, Riverside, California 92521, USA. ²Department of Entomology, University of California, Riverside, California 92521, USA.

³Human Health Division, International Centre of Insect Physiology and Ecology (ICIPE), P.O. Box 30772-00100, Nairobi, Kenya.

*These authors contributed equally to this work.

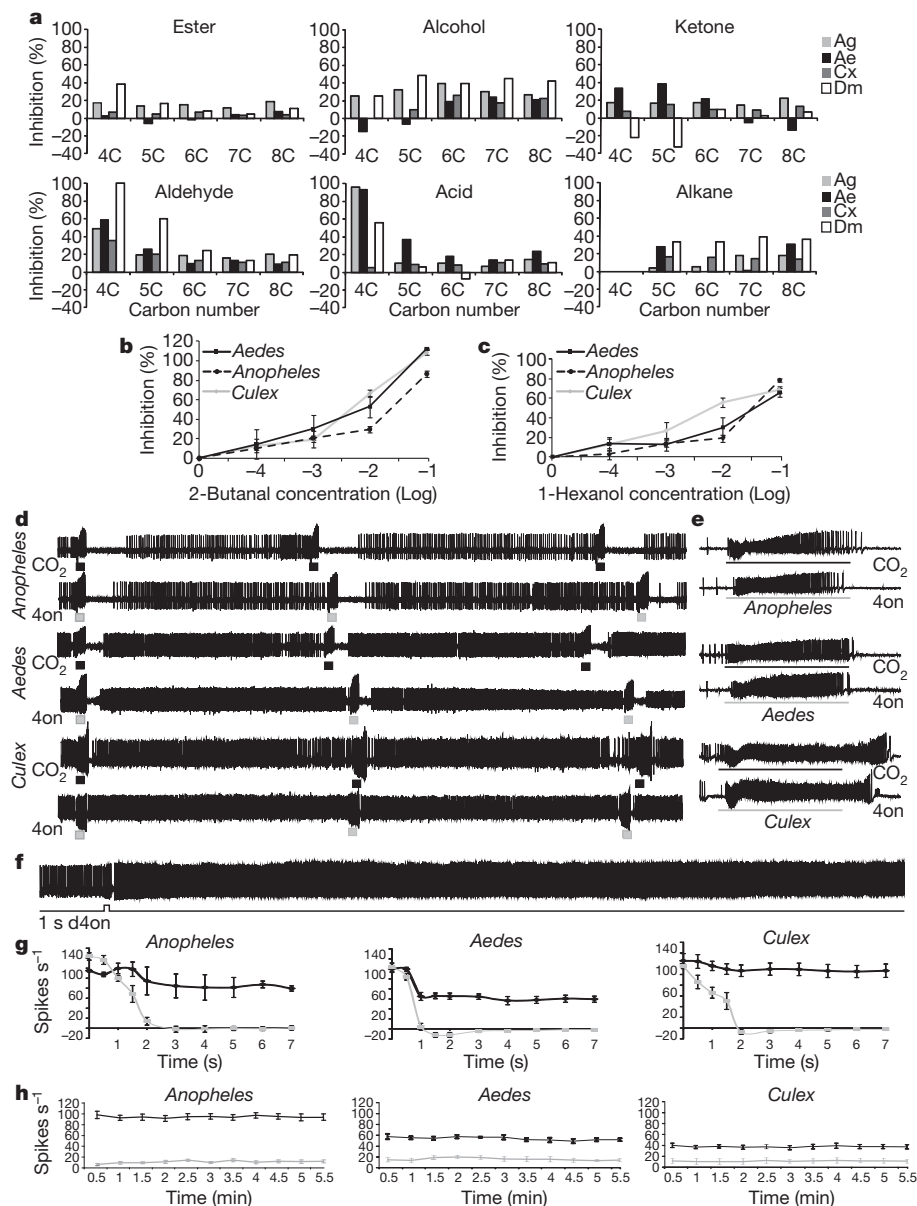


Figure 1 | Identification of odorants that affect the CO₂-sensitive neurons in mosquitoes. **a**, Comparison of percentages of CO₂-response inhibition in cpA neurons of *A. gambiae* (Ag), *A. aegypti* (Ae), *C. quinquefasciatus* (Cx) and *D. melanogaster* (Dm). Functional group is on the primary carbon atom except for ketones on C2, and the length of the carbon chain (carbon number) is indicated. A 1-s stimulus of vapours from odorant diluted 10⁻² applied on cotton wool in cartridge, is overlaid on a 3-s stimulus of 0.15% CO₂. *n* = 3. Data for Dm responses, using 0.33% CO₂, is taken from ref. 19. **b**, **c**, Dose-response of indicated inhibitors in the three species, in a similar overlay assay. *n* = 5; error bars are s.e.m. **d**, Representative long-term recordings from the cp

Most importantly, the brief 2,3-butanedione exposure evoked such a strong prolonged response in cpA neurons of *A. gambiae* and *A. aegypti* that responses to subsequent CO₂ stimuli during the 5.5-min recordings were severely reduced, showing little activation above the high baseline firing of the neuron (Fig. 2a, b). This property is neither enhanced nor reversed efficiently by a subsequent intervening stimulus of another activator or inhibitors (Supplementary Fig. 4b, c), indicating that an active state, perhaps of the receptor, G-protein signalling²⁵, or odorant binding protein may be stabilized by 2,3-butanedione.

To find more effective long-term maskers we tested multiple combinations of CO₂-response-modifying odours and identified a blend of four odorants—2,3-butanedione, 1-hexanol, 1-butanol and 1-pentanol—which is effective at a tenfold lower dilution (10⁻²). This

blend caused prolonged activation and rendered *A. aegypti* cpA neurons unresponsive to CO₂ stimuli throughout the 5.5-min duration of the assay (Fig. 2c, Supplementary Figs 4d, e and 5). The effect of the 'ultra-prolonged' odour(s) is specific to the CO₂-sensitive cpA neuron, as the response of a neighbouring neuron, with a smaller spike amplitude to 1-octen-3-ol, was not affected (Fig. 2d, e). The same ultra-prolonged activating blend was also highly effective on *C. quinquefasciatus* cpA neurons (Fig. 2f).

Aedes females show rapid (<200 ms) and robust responses to entering (turn upwind) and exiting (turn crosswind) a plume of CO₂, both important for source finding². In fact, the cpA neuron shows an exquisite temporal resolution of the presence or absence of above-baseline concentrations of CO₂¹⁸ (Fig. 1d), as required for a system

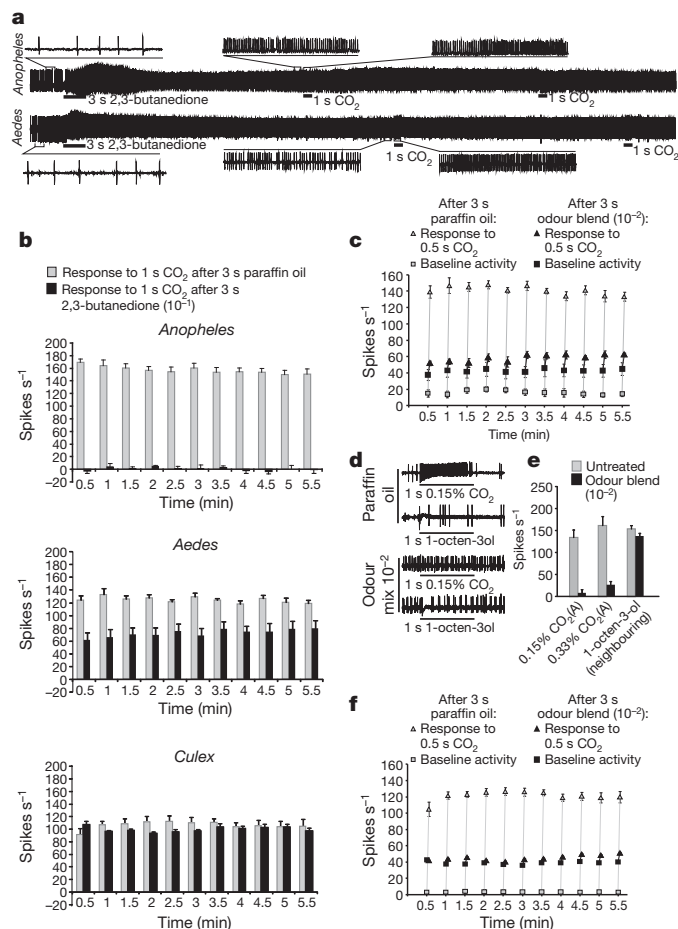


Figure 2 | Ultra-prolonged activation disrupts ability to respond to CO₂. **a**, Representative action potential traces from cp sensillum in *A. gambiae* and *A. aegypti* exposed to a 3-s stimulus of 2,3-butanedione (10^{-1}) followed by two 1-s stimuli of CO₂ (0.15%). Expanded views of the indicated regions of the traces are shown. **b**, Mean increase in frequency of response of cpA neurons to repeated stimuli of 1-s CO₂ (0.15%) applied approximately every 30 s, following a 3-s pre-exposure to 2,3-butanedione (10^{-1} ; black bars) or paraffin oil (grey bars) in *A. gambiae*, *A. aegypti* and *C. quinquefasciatus*. **c**, Mean baseline activity (squares) and increase in activity (triangles) of an *A. aegypti* cpA neuron responding to 1-s CO₂ (0.15%) pulses applied approximately every 30 s after a 3-s pre-exposure to a four-odour blend of 2,3-butanedione, 1-butanol, 1-pentanol and 1-hexanol (10^{-2}) (black) or paraffin oil (grey). **d**, **e**, Representative traces from cp sensillum (**d**) and mean responses from cpA neurons (large spike amplitude) to CO₂ (0.15%) and a neighbouring neuron (small spike amplitude) to 1-octen-3-ol (10^{-3}) after a pre-treatment of 3 min of paraffin oil or the four-odour blend (**e**). **f**, Responses of a cpA neuron in *C. quinquefasciatus* treated identically as in **c**. For **b**, **c**, **e** and **f**, $n = 5$; error bars are s.e.m.

whose sensory input has a key role in the control of flight behaviour. This form of rapid odour-evoked flight modulation (turning and surging) in insects^{26,27} involves both odour detection and discrimination dependent on moment-to-moment changes in olfactory neuron activity²⁷. Higher-order projection neuron circuits are also particularly sensitive to the onset of odour stimuli²⁸. Thus, an ultra-prolonged activator that compromises the ability of sensory neurons to detect rapid changes in CO₂ concentration could disrupt CO₂-mediated attraction behaviour very effectively.

To test the effect of ultra-prolonged activators on behavioural attraction to CO₂, mated *A. aegypti* females were pre-treated with odorant(s) vapours and tested in a wind tunnel (Fig. 3a, b). Mock-treated *A. aegypti* females that were pre-exposed to solvent alone all left the holding cage, crossed the half-way point of the tunnel (Fig. 3c) and navigated successfully along a turbulent CO₂ plume and flew through the CO₂-emitting ring (Fig. 3d). In contrast, females pre-exposed to

the ultra-prolonged odour blend were impaired in their ability to fly upwind (Fig. 3c) as well as in finding the CO₂ source, in a dose- and time-dependent manner (Fig. 3d–f and Supplementary Videos 1–3).

After pre-treatment with the odour blend for 1 min, not only did fewer mosquitoes reach the source (Fig. 3g, h), they also took longer to get there (Supplementary Fig. 6a–c). The proportion of pre-treated mosquitoes that reached the CO₂ source increased as the assay progressed to 5 min (Fig. 3g, h), indicating that their ability to detect CO₂ somewhat improves

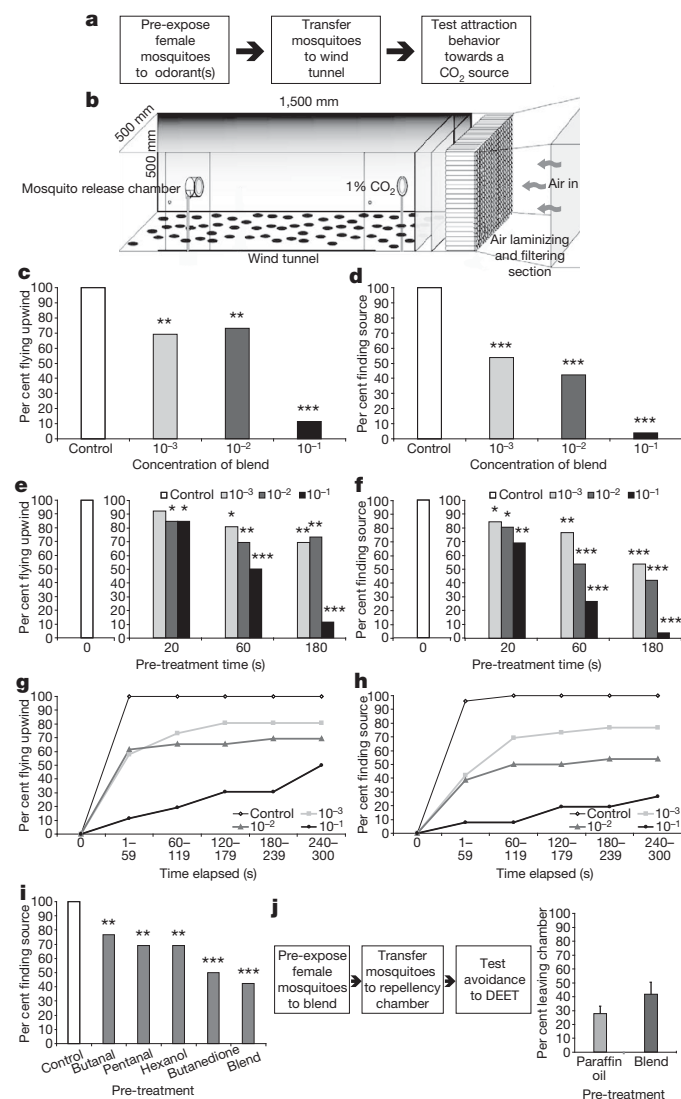


Figure 3 | Exposure to ultra-prolonged activator causes long-term disruption of CO₂-mediated attraction behaviour of female *Aedes* mosquitoes. **a**, **b**, Schematics of the experimental strategy (**a**) and the wind-tunnel apparatus (**b**). **c**, **d**, Percentage of female mosquitoes flying upwind to half-way point (**c**) and flying through the CO₂-emitting glass ring (**d**) after pre-exposure for 3 min to paraffin oil (control) or ultra-prolonged activating blend at indicated concentrations. **e**, **f**, Similar experiment as above except three pre-exposure times were tested to the four-odour ultra-prolonged blend (10^{-1}). **g**, **h**, Time-course of percentages of female mosquitoes getting to half-way point and reaching the CO₂ source after 1 min pre-exposure to paraffin oil or ultra-prolonged blend at the indicated concentrations. **i**, Percentage of female mosquitoes flying through the CO₂-emitting glass ring after pre-exposure to 3 min of paraffin oil (control), indicated odorants (10^{-2}) or ultra-prolonged activating blend (10^{-2}). $N = 26$ individuals for each condition. Pearson's χ^2 test, compared to controls, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. **j**, Schematic of experimental strategy, and percentage of *A. aegypti* mosquitoes escaping repellency chamber within 5 min in response to DEET (9.8%) after pre-exposure to paraffin oil (control) or ultra-prolonged activating blend (10^{-2}) as above. $N = 7$ trials (20 mosquitoes per trial).

with time. Pre-exposure to 2,3-butanedione caused more pronounced behavioural defects in CO₂ attraction than other components of the blend (Fig. 3i).

We determined the specificity of the effect of the odour blend by examining its effect on a different odour-mediated behaviour, and found that avoidance to DEET in a non-contact repellency chamber²⁹ was not reduced by pre-exposure to this blend (Fig. 3j and Supplementary Fig. 7). This result also indicates that, if desired, a long-range CO₂-masking strategy might be used effectively in combination with short-range repellency to DEET.

We then tested for disruption of mosquito hut entry behaviour in a semi-field environment in Kenya³⁰. Host-seeking *C. quinquefasciatus* females were released overnight in a large enclosed greenhouse that contained two faux huts with CO₂-emitting counter-flow geometry traps placed in each of them; females were attracted by the traps and entered the huts through the eaves, and a large majority were then trapped (Fig. 4a and Supplementary Fig. 8). Inclusion of a source of the ultra-prolonged activating blend at 1% (dissolved in paraffin oil) in the form of small fan-driven odour dispensers (Fig. 4b and Supplementary Fig. 8) resulted in a significant reduction in the number of females that

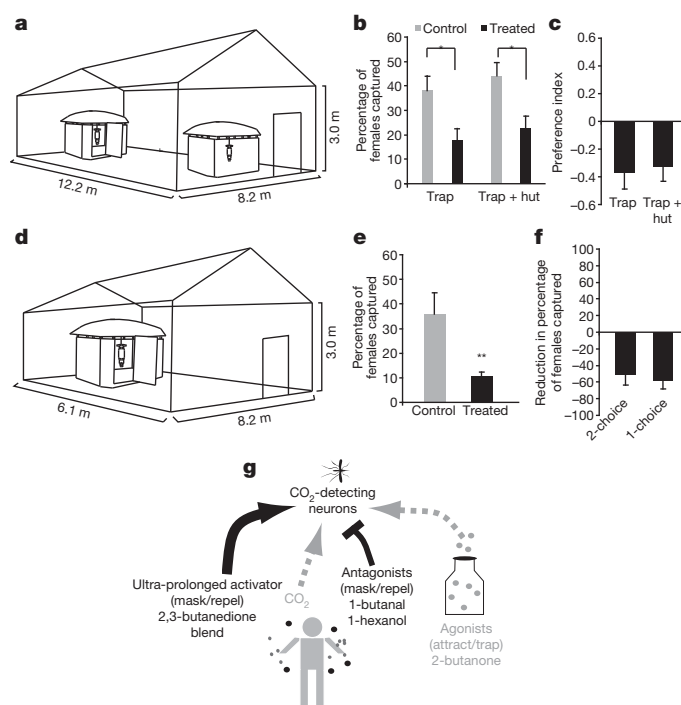


Figure 4 | Ultra-prolonged activators disrupt attraction behaviour of female *Culex* mosquitoes in semi-field conditions **a**, Schematic of two-choice MalariaSphere experiment with counter-flow CO₂ traps placed inside each experimental hut to attract female *Culex* mosquitoes released from the centre of enclosure. One hut contains ultra-prolonged blend dispensers (1%, treated) while the other dispenses paraffin oil (control). **b**, **c**, Mean percentage of released mosquitoes (**b**) and mean preference index of female mosquitoes captured in trap and captured in trap plus hut in treated versus control huts (**c**). $N = 4$ trials, 100 females per trial; error bars are s.e.m.; Student's t -test, $*P < 0.05$. **d**, Schematic of one-choice MalariaSphere experiment. Trials were conducted with (treated) or without (control) ultra-prolonged blend (3%) dispensers. **e**, Mean percentage of mosquitoes in MalariaSphere that are trapped in control and treated huts. $N = 4$ trial nights each for control and treatment, 100 females added every evening; error bars are s.e.m.; Student's t -test, $**P < 0.005$. **f**, Mean reduction in percentage of *Culex* females captured in traps from ultra-prolonged blend treated huts in **b** and **d**. **g**, Proposed model of odorant application for mosquito control. Odorants that are antagonists of the CO₂ receptor can be used to mask mosquito attraction, agonists can be used as a lure or attractant in a trap application, and ultra-prolonged activators can be used to block CO₂ detection through persistent activation of the neuron. Small grey dots represent other human odours, and large black dots represent odours that can be applied to disrupt host seeking.

entered the CO₂ trap in the treated hut. Moreover, the total number of mosquitoes entering the treated hut but not trapped was also significantly lower than in the control hut, both indicating that ultra-prolonged activators masked CO₂-mediated attraction to the hut (Fig. 4b, c). Additionally, odour treatment in one hut did not lead to an increase in the number of mosquitoes entering the control hut (44.25 mosquitoes per night), compared to numbers entering a similarly CO₂-baited hut (56 mosquitoes per night) when odour treatment was not applied to either hut.

In a separate greenhouse experiment, mosquitoes were given unrestricted access to a CO₂-baited trap in a hut through an open door, window and eaves (Fig. 4d). There was a significant reduction in capture of female *C. quinquefasciatus* mosquitoes in the trap when the ultra-prolonged activating blend (3% in paraffin oil) was also continuously dispensed inside the hut (Fig. 4e) as compared to control experiments in the untreated hut. Thus, the odorant blend reduced the number of female mosquitoes inside a hut in both two-choice and one-choice semi-field behaviour assays (Fig. 4f).

Taken together, we show that several aspects of CO₂-mediated behaviour, including activation of upwind flight and navigation towards the CO₂ source, are severely reduced when detection of CO₂ by the cpA neuron is disrupted by exposure to an ultra-prolonged activating blend at concentrations as low as 0.1%. Importantly, the ultra-prolonged blend is absent from the behaviour arena in the wind-tunnel assays. This implies that the reduced attraction towards the CO₂ source cannot be attributed to detection of blend components via other ORNs but is due to the inability of the cpA neurons in pre-treated mosquitoes to detect the CO₂ source.

To test whether an ultra-prolonged activator like 2,3-butanedione can become attractive at lower concentrations (10^{-3} and 10^{-4}), we tested it overnight as an evaporating lure in a counterflow-geometry trap for *Culex* females in a large enclosed arena in California as well as on female *Aedes* mosquitoes for attraction towards a turbulent plume of 2,3-butanedione at 10^{-3} in a wind tunnel. Lower concentrations of the ultra-prolonged activator did not act as an attractive cue for mosquitoes in either assay (Supplementary Table 1).

We describe a novel coding mechanism where ultra-prolonged activation of a sensory neuron modifies its ability to detect and transmit information about changes in concentrations of its ligands in the environment. As well as this class of ligands we have also identified inhibitors and activators of mosquito CO₂ neurons (Fig. 4g). Several characteristics of these novel classes of volatile chemicals offer powerful advantages as potential tools for reducing mosquito-human contact. First, volatile odorants that mask or repel mosquitoes at low concentrations and that can be easily dispersed in the air may protect several individuals within a large area. Second, area masking agents and repellents may be effective without application on skin, and thus the convenience may lead to widespread adoption in developing countries. Third, the use of multiple-odour blends to modify behaviour might delay the emergence of resistant strains, a common concern with other methods of mosquito control. Compounds reported in this initial proof-of-principle investigation such as 2,3-butanedione have undesirable safety profiles at high concentrations and are not ideal for human use without further testing. However, the structures of the agonists and antagonists we have described in this study have enabled us to identify important structural features and pharmacophores (Supplementary Fig. 9) that provide a rational foundation to find additional odorants with improved efficacy and safety profiles. It will be a high priority to identify molecules that are effective at lower concentrations, are environmentally friendly, economical and useful in preventing mosquito-borne diseases.

METHODS SUMMARY

Electrophysiology. Extracellular single-unit recordings were performed as described previously¹⁹ with a few modifications noted in Methods. The odour delivery system was specifically designed to ensure steady levels of CO₂ concentration as described

previously¹⁹ and minimize changes in airflow over insect preparation during single-odour stimulus by switching stimulus airflow from a blank delivery cartridge to an odour-laden delivery cartridge.

Behaviour. Wind-tunnel behaviour experiments were performed as described previously² with some modifications. Avoidance assays with 9.8% DEET were performed in the dark using a repellency chamber that was designed as in ref. 29 with some modifications. Briefly, 20 female *A. aegypti* mosquitoes between 1–2 weeks of age were released inside the repellency chamber and allowed 5 min to ‘escape’ from a spout in response to a 9.8% DEET-impregnated filter paper that was placed out of range of physical contact. Greenhouse behaviour trials were conducted in two MalariaSphere greenhouses³⁰ at the ICIPE field station in Mbita Point, western Kenya. Trapping experiments were also conducted in a netted enclosure inside a greenhouse in the Agricultural Research Station, University of California, Riverside. Details of all behaviour experiments are noted in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 March 2010; accepted 1 April 2011.

- Gillies, M. T. The role of carbon dioxide in host-finding by mosquitoes: a review. *Bull. Entomol. Res.* **70**, 525–532 (1980).
- Dekker, T., Geier, M. & Carde, R. T. Carbon dioxide instantly sensitizes female yellow fever mosquitoes to human skin odours. *J. Exp. Biol.* **208**, 2963–2972 (2005).
- Zwiebel, L. J. & Takken, W. Olfactory regulation of mosquito-host interactions. *Insect Biochem. Mol. Biol.* **34**, 645–652 (2004).
- Ditzen, M., Pellegrino, M. & Vosshall, L. B. Insect odorant receptors are molecular targets of the insect repellent DEET. *Science* **319**, 1838–1842 (2008).
- Syed, Z. & Leal, W. S. Mosquitoes smell and avoid the insect repellent DEET. *Proc. Natl Acad. Sci. USA* **105**, 13598–13603 (2008).
- Krajick, K. Keeping the bugs at bay. *Science* **313**, 36–38 (2006).
- Corbel, V. *et al.* Evidence for inhibition of cholinesterases in insect and mammalian nervous systems by the insect repellent DEET. *BMC Biol.* **7**, 47 (2009).
- Reeder, N. L., Ganz, P. J., Carlson, J. R. & Saunders, C. W. Isolation of a DEET-insensitive mutant of *Drosophila melanogaster* (Diptera: Drosophilidae). *J. Econ. Entomol.* **94**, 1584–1588 (2001).
- Klun, J. A. *et al.* Comparative resistance of *Anopheles albimanus* and *Aedes aegypti* to *N,N*-diethyl-3-methylbenzamide (Deet) and 2-methylpiperidinyl-3-cyclohexen-1-carboxamide (AI3-37220) in laboratory human volunteer repellent assays. *J. Med. Entomol.* **41**, 418–422 (2004).
- Stanczyk, N. M., Brookfield, J. F., Ignell, R., Logan, J. G. & Field, L. M. Behavioral insensitivity to DEET in *Aedes aegypti* is a genetically determined trait residing in changes in sensillum function. *Proc. Natl Acad. Sci. USA* **107**, 8575–8580 (2010).
- Grant, A. J. & O’Connell, R. J. Electrophysiological responses from receptor neurons in mosquito maxillary palp sensilla. *Ciba Found. Symp.* **200**, 233–253 (1996).
- Zollner, G. E., Torr, S. J., Ammann, C. & Meixner, F. X. Dispersion of carbon dioxide plumes in African woodland: implications for host-finding by tsetse flies. *Physiol. Entomol.* **29**, 381–394 (2004).
- Carde, R. T. & Gibson, G. in *Olfaction in Vector-Host Interactions* Vol. 2 (eds Takken, W. & Knols, B. G. F.) 115–141 (Wageningen Academic Publishers, 2010).
- Robertson, H. M. & Kent, L. B. Evolution of the gene lineage encoding the carbon dioxide receptor in insects. *J. Insect Sci.* **9**, 1–14 (2009).
- Lu, T. *et al.* Odor coding in the maxillary palp of the malaria vector mosquito *Anopheles gambiae*. *Curr. Biol.* **17**, 1533–1544 (2007).
- Jones, W. D., Cayirlioglu, P., Kadow, I. G. & Vosshall, L. B. Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila*. *Nature* **445**, 86–90 (2007).
- Kwon, J. Y., Dahanukar, A., Weiss, L. A. & Carlson, J. R. The molecular basis of CO₂ reception in *Drosophila*. *Proc. Natl Acad. Sci. USA* **104**, 3574–3578 (2007).
- Syed, Z. & Leal, W. S. Maxillary palps are broad spectrum odorant detectors in *Culex quinquefasciatus*. *Chem. Senses* **32**, 727–738 (2007).
- Turner, S. L. & Ray, A. Modification of CO₂ avoidance behaviour in *Drosophila* by inhibitory odorants. *Nature* **461**, 277–281 (2009).
- Zdobnov, E. M. *et al.* Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* **298**, 149–159 (2002).
- de Bruyne, M., Foster, K. & Carlson, J. Odor coding in the *Drosophila* antenna. *Neuron* **30**, 537–552 (2001).
- Hall, E. A. & Carlson, J. R. Coding of odors by a receptor repertoire. *Cell* **125**, 143–160 (2006).
- Rostellin, T., Borg-Karlson, A. K., Faldt, J., Jacobsson, U. & Mustaparta, H. The plant sesquiterpene germacrene D specifically activates a major type of antennal receptor neuron of the tobacco budworm moth *Heliothis virescens*. *Chem. Senses* **25**, 141–148 (2000).
- Kaissling, K.-E., Meng, L. Z. & Bestmann, H.-J. Responses of bombykol receptor cells to (Z,E)-4,6-hexadecadiene and linalool. *J. Comp. Physiol. A* **165**, 147–154 (1989).
- Yao, C. A. & Carlson, J. R. Role of G-proteins in odor-sensing and CO₂-sensing neurons in *Drosophila*. *J. Neurosci.* **30**, 4562–4572 (2010).
- Mafra-Neto, A. & Carde, R. T. Fine-scale structure of pheromone plumes modulates upwind orientation of flying moths. *Nature* **369**, 142–144 (1994).
- Bhandawat, V., Maimon, G., Dickinson, M. H. & Wilson, R. I. Olfactory modulation of flight in *Drosophila* is sensitive, selective and rapid. *J. Exp. Biol.* **213**, 3625–3635 (2010).
- Bhandawat, V., Olsen, S. R., Gouwens, N. W., Schlieff, M. L. & Wilson, R. I. Sensory processing in the *Drosophila* antennal lobe increases reliability and separability of ensemble odor representations. *Nature Neurosci.* **10**, 1474–1482 (2007).
- Noosidum, A., Prabaripai, A., Chareonviriyaphap, T. & Chandrapatya, A. Excitatory repellency properties of essential oils from *Melaleuca leucadendron* L. *Litsea cubeba* (Lour.) Persoon, and *Litsea salicifolia* (Nees) on *Aedes aegypti* (L.) mosquitoes. *J. Vector Ecol.* **33**, 305–312 (2008).
- Njiru, B. N., Mukabana, W. R., Takken, W. & Knols, B. G. Trapping of the malaria vector *Anopheles gambiae* with odour-baited MM-X traps in semi-field conditions in western Kenya. *Malar. J.* **5**, 39 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Lacey and S. McNally for help setting up behaviour experiments; the Malaria Group team members in ICIPE Mbita Point Research Station, Y. Afrane and G. Yan for logistical support; S. M. Boyle for help with Pharmacophores; K. Klingler for help with statistics; J. Perecko for building traps and excitorepellency chambers; A. Dahanukar and G. Tauxe for comments on the manuscript; W. Walton, P. Atkinson, P. Wirth and A. Khalon for mosquitoes. Part of this work was funded by a grant to A. Ray from the Bill & Melinda Gates Foundation through the Grand Challenges Exploration Initiative, and part supported by a grant to A. Ray, Award Number R01AI087785, from the NIAID (NIH). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIAID or NIH.

Author Contributions S.L.T. planned the electrophysiology experiments, performed the experiments, collected and analysed the data and helped write the paper. N.L. performed the wind tunnel, repellency, MalariaSphere and greenhouse behaviour experiments and analysed the data. T.G. performed one-choice behaviour experiments. J.G. helped plan MalariaSphere experiments. R.T.C. helped plan the behaviour experiments, and helped write the paper. A.R. helped plan the experiments, analysed the data, managed the project and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.R. (anand.ray@ucr.edu).

METHODS

Electrophysiology. Extracellular single-unit recordings were performed as described previously¹⁹ with few modifications. Chemicals were of the highest purity available, typically >99% (Sigma-Aldrich). Odorants were diluted in paraffin oil at indicated concentrations. Unless indicated, 50 µl of diluted odorant is applied per cartridge, and each cartridge used for no more than three stimuli. A controlled volume of air 5 ml s⁻¹ was puffed through the odour cartridge containing vapours, and was delivered into a constant humidified airstream of 10 ml s⁻¹ that flowed over the fly antenna. The odorant vapour present in the cartridge was thus diluted ~3-fold before being passed over the fly. The CO₂ stimulus was pulsed through a separate delivery system that delivered controlled pulses using a PSM 8000 microinjector (variable 2.5 ml s⁻¹ to 6.5 ml s⁻¹) into the same humidified airstream, from either a 1% or 5% tank of CO₂ (Airgas). The baseline constant humidified airstream (10 ml s⁻¹) was generated from a purified air tank (Airgas) and mixed with a constant controlled volume (5 ml s⁻¹) of filtered room air (~0.035% CO₂). For delivery of binary mixtures of CO₂ with another odorant, we ensured a steady concentration of CO₂ to the fly preparation as described in detail¹⁹. Unless mentioned, responses were quantified by subtraction of the baseline activity immediately preceding stimulus application from the activity during the stimulus. Odour stimuli without CO₂ were also delivered using the system described above to minimize changes in airflow over insect preparation by switching filtered room air (5 ml s⁻¹) from a blank delivery cartridge to an odour-laden delivery cartridge for the indicated duration using the CS-55 stimulus delivery system (Syntech). For each odorant or blend that had a long-term effect on CO₂ response, each recording was obtained from a naive insect.

Behaviour in wind tunnel. Wind-tunnel behaviour experiments were performed as described previously² with some modifications. Briefly, female mated non-blood-fed *A. aegypti* mosquitoes between 7–14 days of age raised in a 14:10 h light:dark cycle were individually collected in release cages and held for 21 h at 25 °C and 70% relative humidity. Mosquitoes in the cage were then transferred into an upended 1-litre glass beaker with 100 µl of odour diluted in paraffin oil at the indicated concentration. The mosquito in the release cage was removed from the odour and beaker and transferred in room air across the room to the wind tunnel (within ~15 s) where it was carefully manipulated into the release cage. Air from outside the building was presented in the wind tunnel in a laminar flow at a controlled rate of 30 cm s⁻¹, at ~70% relative humidity, temperature ~23 °C. A turbulent plume of 1% CO₂ was generated by mixing purified CO₂ and air from cylinders (Airgas) and delivered through a glass ring with eight outlets pointing inwards. The cage was opened remotely to release an individual female mosquito. Two video cameras were used to record the flight path. Analyses of videos were performed on computers offline.

Behaviour in repellency chamber. A stainless steel repellency chamber used a standard design²⁹ (Supplementary Fig. 7). A stainless steel plate was added underneath the smaller screened inner cage to hold the stimulus-loaded paper thus preventing contact with mosquitoes. A blotting paper (12 × 7 inches) was impregnated with 5 ml of 9.8% DEET (Jungle Juice 98% DEET, REI diluted in acetone) and the solvent allowed to evaporate inside a fume hood for 30 min before

placement on the steel plate inside the repellency chamber. Twenty female *A. aegypti* between 1–2 weeks old were transferred to a release cage 30 min before assay start, and pre-exposed to odorant blend exactly as described previously right before start of assay. Mosquitoes were released inside the inner chamber from the release cage and the lights turned off, and the number of mosquitoes escaping from the exit spout at the end of 5 min counted. Equipment was cleaned with acetone or ethanol and dried to avoid contamination.

Two-choice behaviour in MalariaSphere. Experiments were conducted at the Thomas Odhiambo Field Station of the International Centre of Insect Physiology and Ecology in Mbita Point, western Kenya in late April to June. *C. quinquefasciatus* were obtained from the existing rabbit-blood-fed colonies at the station insectary. At 18:00 h 100 female and 20 male *Culex* were released remotely for every experiment from the centre of a large glass-covered MalariaSphere and allowed to enter either of two huts on two sides until 07:00 (Fig. 4a), each running one CO₂-baited (250 ml min⁻¹) counterflow geometry trap (Supplementary Fig. 8). Each hut also ran two TimeMist Fan dispensers starting 30 min before mosquito release and run for the duration of the assay, the test with a container containing 15 ml of odour (1% in paraffin oil), and the control with container of 15 ml paraffin oil. The huts were fitted with one-way entry ports in the eaves as the only mosquito entry points to enable us to count all entering mosquitoes, with windows covered by nets and doors closed. Additionally, huts were lined with clean black cotton sheets that could be removed and washed to reduce odour contamination between trials. Treatment huts and trap positions were randomized. At the end of the trial mosquitoes were counted in each trap and inside each hut. Additional non-participating mosquitoes inside the MalariaSphere were removed.

One-choice behaviour in MalariaSphere. A smaller single-hut net-covered MalariaSphere (Fig. 4d) was used. At 18:00 h, 100 female and 20 male *Culex* mosquitoes were released inside the MalariaSphere and allowed to enter the hut freely through the open eaves, door and window. A CO₂-baited (250 ml min⁻¹) counterflow geometry trap was started at 18:00 inside the hut and trapped mosquitoes were counted at 07:00. One TimeMist Fan dispenser ran continuously during the 4-day duration of the 4 overnight trials with a container containing 15 ml of odour (3%) / paraffin oil which was replaced every 24 h. Because we observed possible long-term contamination effects in thatched roofs and netting in pilot experiments, the no-odour controls were run before odour treatment to avoid contamination. Untrapped mosquitoes were not removed from the MalariaSphere and their approximate numbers taken into account when calculating the percentage caught on each night.

Trapping responses in greenhouse. A greenhouse at the Agricultural research station of University of California, Riverside, was modified to enclose a 3 × 6 m area in netting where 50 laboratory-reared female *C. quinquefasciatus* were released each evening and counterflow geometry traps were run from 17:00 to 09:00 h. CO₂ was dispensed as described earlier. Odorants were dissolved in paraffin oil and dispensed from a 50 ml uncapped tube containing 6 perforations and two 15-cm nylon wicks, which were attached to the airflow outlet tube of the trap. The odorant had largely evaporated from the container at the end of each trial as observed from a sniff test and from change in colour of solution from light yellow to transparent.

Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme

Eric J. Hayden^{1,2}, Evandro Ferrada^{1,2} & Andreas Wagner^{1,2,3}

Cryptic variation is caused by the robustness of phenotypes to mutations¹. Cryptic variation has no effect on phenotypes in a given genetic or environmental background, but it can have effects after mutations or environmental change^{2–5}. Because evolutionary adaptation by natural selection requires phenotypic variation, phenotypically revealed cryptic genetic variation may facilitate evolutionary adaptation^{6–8}. This is possible if the cryptic variation happens to be pre-adapted, or “exapted”⁹, to a new environment, and is thus advantageous once revealed. However, this facilitating role for cryptic variation has not been proven, partly because most pertinent work focuses on complex phenotypes of whole organisms whose genetic basis is incompletely understood. Here we show that populations of RNA enzymes with accumulated cryptic variation adapt more rapidly to a new substrate than a population without cryptic variation. A detailed analysis of our evolving RNA populations in genotype space shows that cryptic variation allows a population to explore new genotypes that become adaptive only in a new environment. Our observations show that cryptic variation contains new genotypes pre-adapted to a changed environment. Our results highlight the positive role that robustness and epistasis can have in adaptive evolution^{10,11}.

For our evolution experiments, we chose the well-characterized group I RNA enzyme (ribozyme) derived from the *Azoarcus* pre-tRNA^{Leu}. This ribozyme retains activity at unusually high temperatures (80 °C), or in the presence of high concentrations of denaturants (7.5 M urea)¹². Thus, we expected this robust phenotype to tolerate mutations without losing function, making it an ideal candidate for the experimental study of cryptic variation. In addition, detailed functional and structural data on the *Azoarcus* ribozyme provide guidance for designing and interpreting *in vitro* evolution experiments^{13–16}. We used an *in vitro* procedure for the evolution of a group I intron^{17,18}. In this procedure, populations of RNA molecules evolve on the basis of Darwinian principles, through cycles of random mutagenesis and selection based on the preservation of sequences that successfully perform a catalytic task (Supplementary Fig. 1).

To produce cryptic variation, we exposed ribozyme populations to mutagenesis followed by purifying selection for the ‘native’ activity of RNA phosphate bond cleavage (Supplementary Fig. 1b). We introduced mutations at an average rate of one mutation per individual per generation (Methods). We expected that evolving populations would accumulate mutations while maintaining the native function. We carried out two independent evolution experiments, and called them “line A” and “line B”. They were identical except for the addition of a denaturant (5 M formamide) to line B, which in terms of structural stability is analogous to increasing the temperature by 13.5 °C (ref. 19; Supplementary Fig. 2).

We monitored the native activity of the two lines over ten generations of purifying selection while introducing mutations at every generation (Fig. 1a). No significant difference in activity (*t*-test, 95% confidence interval) exists between the initial and final round of selection for both lines. This demonstrates that the procedure was maintaining the native activity of the population despite mutagenesis, and supports the cryptic nature of any accumulated mutations.

We confirmed mutation accumulation through DNA sequencing of 2,748 ± 770 molecules sampled from each generation (Methods). We determined the mutational distance between individuals in the populations and the original wild-type sequence, which confirmed an increase in population distance from the wild-type sequence over time (Fig. 1b, c). By generation 10 less than 1% of the sampled individuals had no mutations (Supplementary Fig. 3). Mutation accumulation varied at each position in the ribozyme sequence; 35 positions in line A and 19 positions in line B were “mutable”, with a rate of accumulation significantly greater than zero (Methods). Of these positions, 15 were common to both lines.

Although the accumulated mutations did not affect the phenotype of the population under ‘native’ conditions, we proposed that they could facilitate evolutionary adaptation to a new chemical environment.

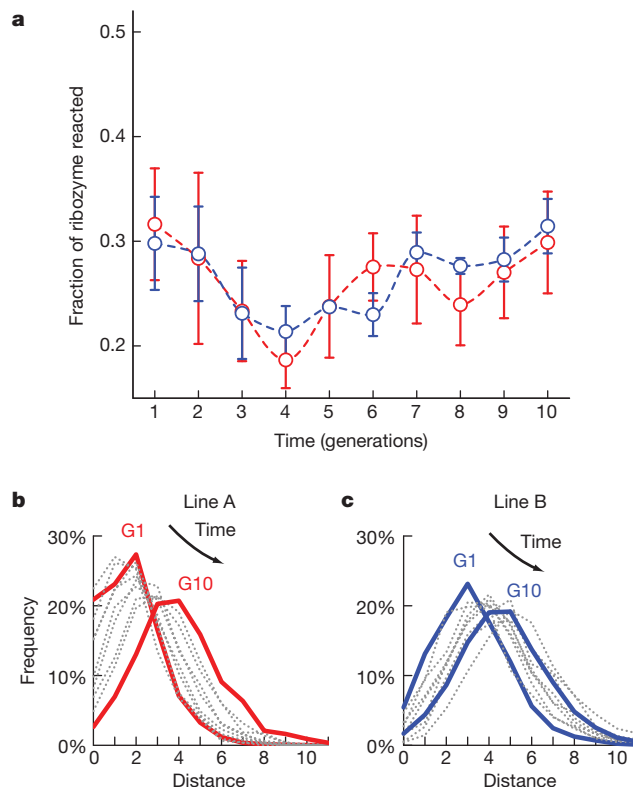


Figure 1 | Evolution during selection for the native activity. **a**, Activity (fraction of ribozyme reacted) at each generation under conditions used during selection for the native activity (RNA oligonucleotide cleavage) over 10 generations (Methods). Error bars correspond to standard errors of three measurements. **b**, **c**, Histograms, from each generation of line A and B, showing the frequency (percent of sample) of individuals with a given number of nucleotide differences from the wild-type sequence (distance). Frequencies from generations 1 (G1) and 10 (G10) are shown as solid lines, and intervening generations are shown as dotted grey lines.

¹Institute of Evolutionary Biology and Environmental Studies, University of Zurich, 8057 Zurich, Switzerland. ²The Swiss Institute of Bioinformatics, Quartier Sorge-Batiment Genopode, 1015 Lausanne, Switzerland. ³The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA.

Thus, we challenged the resulting populations to adapt to a non-native function by changing the substrate in the selection procedure. We chose an RNA substrate with identical sequence, but with a phosphorothioate replacing the scissile phosphate. This chemical change represents a “promiscuous activity”²⁰ of the *Azoarcus* ribozyme with a ~200% decreased catalytic efficiency (k_{cat}/K_m), by mostly affecting k_{cat} (ref. 13). We started new evolution lines from $\sim 10^{13}$ (20 pmol) RNA molecules taken from the last generation of line A and line B. We called these lines New-A and New-B, respectively. In addition, we started another new line from a sample of $\sim 10^{13}$ RNA molecules taken from the original initial population. We called this line New-WT. In this phase of the experiment we wanted to analyse the effect of previously accumulated mutations on evolutionary adaptation to the new substrate. Thus, we used the same reaction conditions for all three lines, and reduced the mutation rate to ~ 0.16 per individual per generation, by replacing the mutagenic PCR step of our selection procedure with a standard PCR (Methods).

We selected for activity with the new substrate during eight generations, and measured the activity of each population at each generation (Fig. 2a). In each line the activity increased significantly between the first and last generations. However, lines New-A and New-B showed a much faster rate of adaptation than line New-WT. We calculated the rate of adaptation by dividing the percent increase in the fraction of ribozyme reacted by time (generations). The greatest difference in rate is found at generation 5, where the rates of adaptation for lines New-A, New-B and New-WT were 19.5, 15.5, and 2.5, respectively, corresponding to an ~eightfold faster rate of evolutionary adaptation for line New-A, and an ~sixfold faster rate for line New-B, relative to line New-WT.

Next we identified genotypes (combinations of mutations) that were potentially contributing to the increasing activities of the evolving populations on the basis of their rapid frequency increase within the population (Methods). Two important genotypes stand out (Fig. 2b). In line New-A, the most rapidly increasing genotype, termed *AzoΔ*, includes deletions at positions 47–53 combined with seven point mutations (G31U, G35U, G70U, G121A, C141U, A144G and G183C). By generation 8, this genotype represented 31% of the population. In line

New-B, the most rapidly increasing genotype, termed *Azo**, is composed of four point mutations (G32U, G53A, C89U and G179C). By generation 8, sequences containing all four mutations accounted for 23% of the population, and various subsets of these four mutations accounted for 78% of the population. All pairs of mutations of each genotype showed significant correlation coefficients²¹ ($P < 0.05$, chi-squared), supporting the conclusion that each combination of mutations usually occurs together in the same individual (Supplementary Table 1).

To confirm the selective advantage of these genotypes, we synthesized clonal transcripts of the *AzoΔ*, *Azo** and wild-type ribozymes for kinetic analysis (Methods). For the *Azo** ribozyme, we found an increased activity with a phosphorothioate substrate as compared to the wild-type *Azoarcus* ribozyme (Fig. 2c). The four mutations of *Azo** increase the observed rate constant (k_{obs}) by 131%, and also increase the extent of ribozyme reacted by 76%. Thus, the presence of this genotype in the population accounts for much of the increasing activity of line New-B.

Surprisingly, the clonal preparation of the *AzoΔ* ribozyme showed no activity towards the phosphorothioate substrate. We proposed that this sequence lacked the ability to fold into the native state individually, but could form an active complex in conjunction with other active ribozymes. Such an intermolecular partnership was observed in several other ribozyme experiments^{22–24}. To test this hypothesis, we assayed the *AzoΔ* ribozyme for activity with a phosphorothioate substrate alone, or in the presence of either the wild-type or the *Azo** ribozymes (Fig. 2d). In these experiments, only the *AzoΔ* ribozyme was 5′-radiolabelled with ³²P, so that only the activity of this ribozyme was observable on a denaturing polyacrylamide gel. The results confirm that although the *AzoΔ* ribozyme is inactive individually, it regains activity upon addition of either active variant.

We also looked for the presence of the highly active *Azo** genotype in lines New-A and New-WT. In line New-A, *Azo** is present and increases in frequency from 1.1% to 8.0% over eight generations (Fig. 2b). The lower fitness of *Azo** in this line is presumably due to the presence of individually inactive, yet highly fit *AzoΔ* variants. Because the increase in the frequency of *Azo** is modest in this line, linear regression had not identified the four mutations as individually significant, which demonstrates a limitation of the regression approach.

Although the *Azo** genotype showed increased activity with the phosphorothioate substrate, this genotype did not rise to high frequency in line New-WT, which had not acquired cryptic variation. This genotype did not appear in the first three generations, and we found only three individuals in generation 8 (0.2% of sample). Analysis of correlation coefficients in line New-WT confirms that *Azo** mutations rarely occur together in the same individual (Supplementary Table 1).

Although the *Azo** genotype has an increased activity in the new environment, our data indicate that the mutations that comprise this genotype had no advantage in the native environment. First, the composite activities of the populations of lines A and B did not increase during selection for the native activity, indicating that these lines had not yet discovered higher fitness genotypes. Also, the individual mutations of the *Azo** genotype do not dramatically increase in frequency during selection for the native activity (Supplementary Figs 4 and 5), and the *Azo** genotype (all four mutations) was not detected in line A or B.

To confirm the cryptic nature of the *Azo** mutations, we engineered these mutations individually into the *Azoarcus* ribozyme. We then determined the activity of these variants under the conditions used during selection for the native activity (no formamide) and compared them to the wild-type activity (Supplementary Fig. 6). Only G179C causes an increased mean activity (14%), but that is not significantly different from the activity of the wild-type ribozyme ($P = 0.11$, *t*-test). mutation C32U causes a significantly decreased activity (–33%,

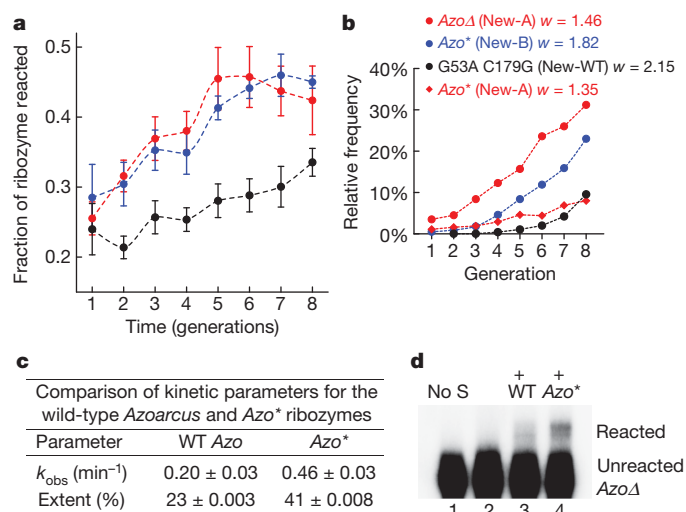


Figure 2 | Evolution during selection for the new activity. **a**, Activity (fraction of ribozyme reacted) at each generation under conditions used during selection for phosphorothioate bond cleavage, with standard error based on three measurements. **b**, Frequency of genotypes (percent of sample) over time (generations), and their corresponding relative fitness w . **c**, Comparison of kinetic parameters for the *Azo** and wild-type ribozymes. **d**, Intermolecular activity of the *AzoΔ* ribozyme, under the same conditions as during selection (Methods): 200 pmol phosphorothioate substrate, 20 pmol 5′-[³²P]-labelled *AzoΔ*. In addition, lanes 3 and 4 contained 40 pmol wild-type and *Azo**, respectively. The negative control ‘No S’ contained no substrate.

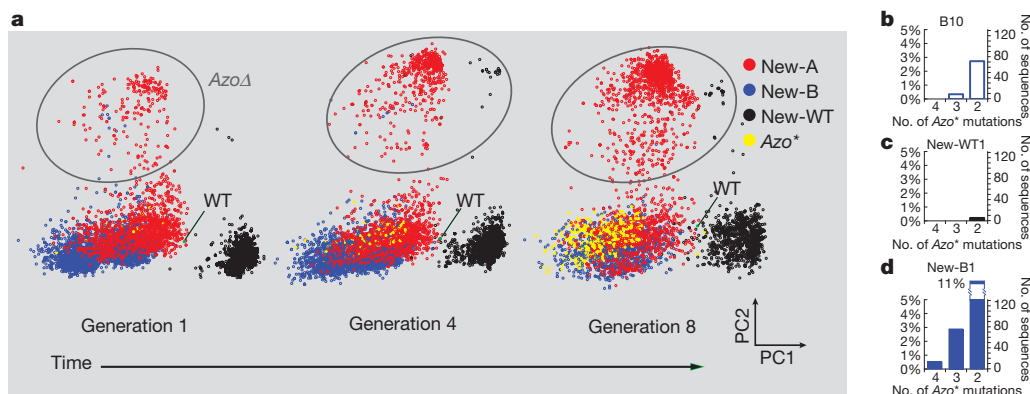


Figure 3 | Evolution in genotype space. **a**, Principal component analysis of pooled sequence data from New-A, New-B and New-WT populations. The first two principal components are shown ('PC1' and 'PC2'). Nodes represent individual sequences. The distance between nodes is proportional to the number of nucleotide differences, but may appear decreased due to the compression of multiple dimensions. The region on the graphs occupied by the

AzoΔ sequences is indicated by a grey ellipse. **b–d**, Frequency of sequences with a given number of the *Azo** mutations in generation 10 of line B (B10, **b**), and the first generation of line New-WT (New-WT1, **c**) and line New-B (New-B1, **d**). Frequencies are presented as percentage (left y-axis) and total number (right y-axis).

$P = 0.03$). mutations G53A and C89U both cause non-significant decreased activities (-28% , $P = 0.07$ and -17% , $P = 0.10$, respectively). We conclude that the individual mutations of the *Azo** genotype had no fitness benefit during selection for the native activity. However, because three of these individual mutations, and several combinations (Supplementary Fig. 6), show no significant difference from the wild type, they can remain in the population despite purifying selection for the native activity. This is consistent with the observations that $\sim 10\%$ of sampled individuals in generation 10 of lines A and B had at least one of the *Azo** mutations, but none of the mutations showed a marked increase in frequency.

Next we turn to a more detailed visual analysis of sequence space to help us understand why cryptic variation allowed faster adaptation. This space is very high-dimensional and cannot be visualized directly. However, we can study lower-dimensional projections of this space using principal component analysis of aligned sequence data sampled from evolving populations. Figure 3 shows such an analysis based on sequences isolated from three generations of the New lines. It shows that, first, lines New-A and New-B are more diverse during all generations, compared to line New-WT. Second, this analysis confirms the existence of two subpopulations of line New-A, where two clearly discernible clouds of sequences are visible at all times, one contains *Azo** and the other contains *AzoΔ*. Third, it illustrates the high fitness of *AzoΔ* and *Azo** in that the number of sequences belonging to these genotypes increases over generational time. Importantly, it shows that many of the sequences in generation 1 of lines New-A and New-B are close in genotype space to *Azo**. Over time, the genotypes become more concentrated around the *Azo** genotype. In contrast, in generation 1 of line New-WT, sequences are tightly clustered and distant from *Azo**. Over time, this population becomes more diverse, and moves towards the region of space occupied by the *Azo** individuals.

A candidate explanation for the advantage of cryptic variation that emerges from the previous analyses is that lines A and B had the opportunity to expand in sequence space, such that their sequences came close to regions where advantageous mutations could occur in line New-A and New-B. Line New-WT did not have this opportunity, and thus adapts more slowly. Thus, although the genetic variation acquired during purifying selection did not affect the population activity on the native substrate, it allowed for rapid adaptation after the environmental perturbation. This rapid adaptation coincides with the rise of *Azo**, a variant with increased activity. The proximity of line B individuals to *Azo**, relative to line New-WT individuals, is supported by analysing the positions in the ribozyme sequence where the *Azo** mutations occur: 32, 53, 89 and 179 (Fig. 3b–d). The results show that many sequences in generation 10 of line B (B10) already possess two or three of

the four *Azo** mutations. No individuals in the first generation of line New-WT (New-WT1) possesses three *Azo** mutations, and only a fraction of a percent possess two. The proximity of line B individuals is also supported by a clustering analysis, which shows that sequences that cluster near *Azo** are present in B10 but not in New-WT1 (Supplementary Fig. 7). Thus, the cryptic diversity acquired during purifying selection for the native activity moved some of the population to regions of genotype space that happen to be proximal to a genotype with high fitness for the New substrate.

Our observations demonstrate that cryptic variation can facilitate adaptation, and why. Populations under purifying selection for a trait can evolve genotypic diversity, if there are many different genotypes with the same or similarly well-adapted phenotype. Some of this diversity is fortuitously pre-adapted to new environments, which aids the population's evolutionary adaptation after environmental change. We note that this genotypic diversity is a signature of extensive epistasis. Indeed, such epistasis has recently been demonstrated in protein and RNA phenotypes^{25–27}. Epistasis is important in our system, because several individual mutations do not provide a large fitness advantage alone, but do so in combination. The ability to explore such combination of mutations cryptically is especially important in cases where high-fitness genotypes require several interacting mutations. These observations support theoretical work which demonstrates that the release of hidden variation after perturbation is a general property of genetic systems near mutation–selection balance with epistatic or gene–environment interactions²⁸.

The phenotype of our study system is much simpler than complex traits of higher organisms. However, with this system we can monitor population-wide genotypic change over multiple generations and study the relationship, in genotype space, between standing variation and high fitness genotypes. Our results suggest that we may understand the role of cryptic variation in complex organismal traits to the extent that we can analyse their evolution in an underlying genotype space.

METHODS SUMMARY

The double-stranded DNA template for the *Azoarcus* ribozyme was produced from a two-step PCR-based assembly of synthetic oligonucleotides²⁹. Ribozyme populations were prepared from *in vitro* transcription (T7 RNA polymerase) and purified for length homogeneity by denaturing PAGE (6% polyacrylamide/8 M urea). Mutagenesis was achieved by a mutagenic PCR procedure²⁴, and to a lesser extent by the inherent mutation rates of the polymerase enzymes of the selection procedure. Substrate oligonucleotides were produced by solid phase synthesis and purified by PAGE (Microsynth). Selection reactions and activity assays contained 20 pmol ribozymes, and either 100 pmol RNA oligonucleotide substrate, or 200 pmol phosphorothioate containing substrate (equal mixture Rp/Sp). Negative controls for the selection protocol were carried out for every generation by skipping

the reverse transcription step, but keeping the remainder of the protocol identical, and were monitored at both PCR steps by agarose gel electrophoresis. No band was ever observed in a negative control. Kinetic parameters were determined by nonlinear curve fitting of time course data (Methods). Complementary DNA samples from each generation were appended with a primer sequence unique to that generation via a PCR reaction. Samples from all generations were combined and sequenced on a single picotitre plate with a GS-FLX system (Roche /454 Life Sciences) at the Functional Genomics Facility, Zurich. *P*-values from linear regression were adjusted for multiple testing using the Benjamini Hochberg procedure³⁰. Principle component analysis was performed using the princomp function in Matlab on multiple sequence alignments of data from pooled generations.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 September 2010; accepted 1 April 2011.

- Wagner, A. *Robustness and Evolvability in Living Systems*. (Princeton University Press, 2005).
- Gibson, G. & Dworkin, I. Uncovering cryptic genetic variation. *Nature Rev. Genet.* **5**, 681–690 (2004).
- Waddington, C. H. Genetic assimilation of an acquired character. *Evolution* **7**, 118–126 (1953).
- Rutherford, S. L. From genotype to phenotype: buffering mechanisms and the storage of genetic information. *Bioessays* **22**, 1095–1105 (2000).
- Le Rouzic, A. & Carlborg, O. Evolutionary potential of hidden genetic variation. *Trends Ecol. Evol.* **23**, 33–37 (2008).
- Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
- True, H. L. & Lindquist, S. L. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **407**, 477–483 (2000).
- Masel, J. Cryptic genetic variation is enriched for potential adaptations. *Genetics* **172**, 1985–1991 (2006).
- Gould, S. J. & Vrba, E. S. Exaptation—a missing term in the science of form. *Paleobiology* **8**, 4–15 (1982).
- Aharoni, A. *et al.* The 'evolvability' of promiscuous protein functions. *Nature Genet.* **37**, 73–76 (2004).
- Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proc. Natl Acad. Sci. USA* **103**, 5869–5874 (2006).
- Tanner, M. & Cech, T. Activity and thermostability of the small self-splicing group I intron in the pre-tRNA(Ile) of the purple bacterium *Azoarcus*. *RNA* **2**, 74–83 (1996).
- Kuo, L. Y., Davidson, L. A. & Pico, S. Characterization of the *Azoarcus* ribozyme: tight binding to guanosine and substrate by an unusually small group I ribozyme. *Biochim. Biophys. Acta* **1489**, 281–292 (1999).
- Stahley, M. R. & Strobel, S. A. Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* **309**, 1587–1590 (2005).
- Rangan, P., Masquida, B., Westhof, E. & Woodson, S. A. Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme. *Proc. Natl Acad. Sci. USA* **100**, 1574–1579 (2003).
- Pan, J. & Woodson, S. A. Folding intermediates of a self-splicing RNA: mispairing of the catalytic core. *J. Mol. Biol.* **280**, 597–609 (1998).
- Beaudry, A. A. & Joyce, G. F. Directed evolution of an RNA enzyme. *Science* **257**, 635–641 (1992).
- Lehman, N. & Joyce, G. F. Evolution *in vitro* of an RNA enzyme with altered metal dependence. *Nature* **361**, 182–185 (1993).
- Blake, R. D. & Delcourt, S. G. Thermodynamic effects of formamide on DNA stability. *Nucleic Acids Res.* **24**, 2095–2103 (1996).
- O'Brien, P. J. & Herschlag, D. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **6**, R91–R105 (1999).
- Hartl, D. L. & Clark, A. G. *Principles of Population Genetics*. (Sinauer Associates, 1997).
- Hanczyc, M. M. & Dorit, R. L. Experimental evolution of complexity: *in vitro* emergence of intermolecular ribozyme interactions. *RNA* **4**, 268–275 (1998).
- Lawrence, M. S. & Bartel, D. P. New ligase-derived RNA polymerase ribozymes. *RNA* **11**, 1173–1180 (2005).
- Ouellet, J., Byrne, M. & Lilley, D. M. Formation of an active site *in trans* by interaction of two complete Varkud Satellite ribozymes. *RNA* **15**, 1822–1826 (2009).
- Kondrashov, F. A. & Kondrashov, A. S. Multidimensional epistasis and the disadvantage of sex. *Proc. Natl Acad. Sci. USA* **98**, 12089–12092 (2001).
- Wilke, C., Lenski, R. & Adami, C. Compensatory mutations cause excess of antagonistic epistasis in RNA secondary structure folding. *BMC Evol. Biol.* **3**, 3 (2003).
- Weinreich, D. M., Delaney, N. F., Depristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
- Hermisson, J. & Wagner, G. P. The population genetic theory of hidden variation and genetic robustness. *Genetics* **168**, 2271–2284 (2004).
- Rydzanicz, R., Zhao, X. S. & Johnson, P. E. Assembly PCR oligo maker: a tool for designing oligodeoxynucleotides for constructing long DNA molecules for RNA production. *Nucleic Acids Res.* **33**, W521–W525 (2005).
- Hochberg, Y. & Benjamini, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Dhar and N. de la Chaux for bioinformatic support. We thank the Functional Genomics Center, Zurich, for collecting sequence data. We acknowledge support from Swiss National Science Foundation grants 315200-116814, 315200-119697 and 315230-129708, from the YeastX program of SystemsX.ch, and from the Research Priority Program in Systems Biology at the University of Zurich.

Author Contributions E.J.H. and A.W. contributed to the design of the experiments; E.J.H. performed the experiments; E.J.H., E.F., and A.W. all contributed to analysis of the data and co-wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.W. (andreas.wagner@ieu.uzh.ch).

METHODS

RNA preparation. The dsDNA templates containing variants of the *Azoarcus* ribozyme were constructed by a two-step PCR-based assembly from six synthetic DNA oligonucleotides²⁹. The templates contain 197 nucleotides of the *Azoarcus* group I intron, excluding the first eight nucleotides but including the nucleophilic terminal guanosine (G205), all preceded by the T7 promoter sequence to allow *in vitro* transcription. Transcription reactions (200 µl) contained 50 mM Tris pH 7.5, 15 mM MgCl₂, 5 mM DTT, 2 mM spermidine, approximately 160 ng dsDNA template, T7 RNA polymerase (unknown concentration), and RNase-free water (Ambion), and were incubated at 37 °C for at least 4 h. Reactions were then DNase treated to remove the DNA template at 37 °C for 30 min with 10 U RNase-free DNase I (Promega). Reactions were stopped with the addition of 15 mM EDTA, and extracted two times with phenol:chloroform (5:1, pH 4.5, Ambion) to remove protein enzymes and remaining DNA template. Reactions were ethanol-precipitated and rehydrated in equal volumes RNase-free water (Ambion) and a formamide loading dye. RNA was purified for length homogeneity by denaturing polyacrylamide gel electrophoresis (PAGE, 6% polyacrylamide, 8 M urea). Purified RNA was visualized by ultraviolet light, excised from the gel, and eluted into 0.3 M sodium acetate (pH 5.5) by diffusion. Eluted RNA was passed through a 0.2 µm spin column filter (VWR), ethanol-precipitated, and rehydrated to a desired concentration in RNase-free water (Ambion). Substrate oligonucleotides GGCAU(AAAU)_sA and GGCAUs(AAAU)_sA (s = phosphorothioate bond) were produced by solid phase synthesis and purified by denaturing PAGE (Microsynth). Concentrations were determined by ultraviolet-absorbance on an ND-1000 spectrophotometer (NanoDrop Technologies).

Mutation rates and diversity calculations. The per generation mutation rates of our selection procedure were estimated on the basis of previously reported mutation rates from a publication that quantified the number of mutations resulting from reverse transcription and PCR of a group I ribozyme³¹. The authors calculated a mutation rate for both a mutagenic PCR procedure and a standard “non-mutagenic” PCR procedure. For the mutagenic PCR procedure, the authors reported a mutation rate of 0.0066 ± 0.0013 per nucleotide per PCR (95% confidence interval, $n = 16,591$), with very low mutational bias, by including the following in the reaction mixture: manganese (0.5 mM), a biased nucleotide pool (5:5:1:1 ratio of dCTP:dTTP:dATP:dGTP), and elevated levels of magnesium (7 mM) as well as *Taq* polymerase (5 U, NEB) relative to a standard PCR. We used these mutagenic PCR conditions to generate the initial populations used to start lines A, B, New-WT, and to introduce mutations at each generation of lines A and B. For the standard PCR protocol, the authors reported a mutation rate of 0.001 per nucleotide per PCR. Our standard PCR conditions were very similar, and we used this mutation rate as an estimation of our mutation rate when these conditions were used instead of the mutagenic PCR, that is, in the New lines. The lower mutation rate is very consistent with other calculations of PCR mutation rates under ‘standard’ conditions that are similar to those used during our New selection lines³². Using the average of the eight reported values of per nucleotide per cycle mutation rates (p), our per PCR mutation rate (f) can be calculated as $f = np/2$, where n is the number of doublings observed in our evolution procedure ($n = 27$, from $\sim 6 \times 10^8$ fold amplification over two PCRs). Using this formula we calculate an expected mutation rate of 0.00116 per nucleotide per PCR for our non-mutagenic procedure.

We calculated the average mutation per individual in our population as the per nucleotide per PCR mutation rate times the length of the mutable region of our ribozyme sequence (inside the primer binding sites). This calculation gives an average number of mutations per individual of 1.05 under mutagenic conditions and 0.159 under our ‘non-mutagenic’ conditions. The expected composition of populations produced from the given error rate was also calculated using binomial statistics³³. The probability P of a molecule having k mutations in a population with a mutagenized region of length l produced with a mutation rate of m was calculated as: $P(k,l,m) = [l!/(l-k)!k!]m^k(1-m)^{l-k}$.

Selection procedure. Active variants were selected based on a reverse splicing reaction containing 20 pmol ribozyme population, 30 mM EPPS (pH 7.5), 25 mM MgCl₂, and either 100 pmol RNA oligonucleotide substrate, or 200 pmol phosphorothioate-bond-containing substrate (equal mixture Rp/Sp). Reactions were incubated at 37 °C for 1 h. A sample of this reaction (20%) was then directly subjected to a reverse transcription reaction containing 1 mM dNTPs, 5 units AMV-RT (Fermentas), the commercially supplied buffer, and 200 pmol of a DNA primer (5'-TATTTATTTATTTATTTCC-3') complementary to the 3' end of the substrate, and to the final two nucleotides of the ribozyme. A portion of the resulting cDNA (5%) was used as a template in a ‘selective’ PCR containing 2 mM dNTPs, 1.5 mM MgCl₂, 10 pmol of the reverse transcription primer, and 10 pmol of a primer complementary to the 5'-end of the ribozyme (5'-CCGGTTTGTGTGACTTTGCC-3'). Approximately 0.1 fmol of the resulting DNA was subjected to a ‘regenerative’ PCR reaction, with primers that restore the 3' end of

the ribozyme sequence, and that append a T7 promoter sequence to the 5' end. In lines A and B, the regenerative PCR was done under mutagenic conditions, as described above. In New lines, it was performed under standard ‘non-mutagenic’ conditions. For all PCR reactions we used a standard *Taq* polymerase (NEB), as well as standard *Taq* buffer: 10 mM Tris, 50 mM KCl, 1.5 mM MgCl₂, pH 8.3 (NEB). Selection-negative controls were conducted to control for amplification of unreacted ribozymes and contaminating DNA by skipping the reverse transcription step, while keeping the rest of the protocol identical. These controls were monitored at both PCR steps by agarose gel electrophoresis. No bands were observed in the negative controls at any generation.

Kinetic experiments. Kinetic experiments were carried out under the same conditions as the selection reactions (above). The *Azoarcus* ribozyme variants were prepared side by side to minimize sample to sample variation. Reactions were performed in 40 µl volumes, and 5 µl aliquots were removed at six time points. The fraction of ribozyme converted to a 3'-modified species by ligation of a portion of the substrate was calculated from the relative fluorescent intensity of bands after separation by denaturing PAGE (6% acrylamide, 1,000 Vh) and staining with GelRed (Biotium). Reactions were carried out in triplicate and fit to the equation $F(t) = A(1 - e^{-kt})$, where F is the fraction reacted at time t , and A and k are non-linear fitting parameters used to estimate the extent of reaction and the observed rate, respectively. Fitting was performed with Gnuplot (www.gnuplot.info).

Ultra high-throughput (UHT) sequencing. Approximately 0.03 pmol of selective PCR product was used as input into an additional PCR to prepare samples for UHT-sequencing. Reactions (50 µl) contained 1.25 units *Taq* polymerase (NEB), standard *Taq* buffer (NEB), 2 mM dNTPs, and 15 pmol primers with extensions that allow compatibility with a GS FLX System (Roche, 454 technology). In addition, one of the primers contained a unique 6 nucleotide sequence. Each unique sequence differed by at least two nucleotides to prevent confusion by sequencing errors. A different unique sequence was used for each generation in the study. PCR products were quantified on agarose gels stained with GelRed (Biotium), and adjusted to approximately equal concentrations. A portion (3 µl) of the resulting PCR products were pooled together, and sequenced on a single picotitre plate. After sequencing, the unique primer sequence was used to sort sequences by line and generation. Sequences shorter than 95% of the wild-type sequence or with average quality scores less than 35 were discarded.

Linear regression and false discovery rate control. A multiple sequence alignment was constructed separately for sequences determined in each population and each generation using the Needleman–Wunsch algorithm³⁴. For each position in each alignment, the frequency of mutation was calculated. The regression coefficient (slope) was determined for each position by linear regression of mutational frequency over generation time. A t -test was performed and raw P -values were calculated for each coefficient to determine significance relative to either zero (lines A and B) or the expected mutation rate in that line (see above). Raw P -values were rank-ordered, and converted to false-discovery-rate-controlled P -values by the Benjamini and Hochberg procedure³⁰.

Genotype identification. We determined the population frequency of different combinations of mutations, and identified genotypes that increased most rapidly in their frequency with respect to generation time. We limited our search to genotypes that were comprised of combinations of mutations that individually showed a significant increase in frequency with time. The significance of increase for individual mutations was determined by linear regression with correction for multiple testing (Supplementary Fig. 4). It should be pointed out that the increase in the frequency of genotypes that are under positive selection is expected to be exponential, not linear, under standard population genetic models²¹. Linear regression penalizes nonlinearity, and thus our determination of significance of individual mutations is a conservative approach.

Mutational co-occurrence. For each pair of mutations deemed significant by linear regression, the frequency that each mutation occurred without the other was determined and termed P_{Ab} and P_{aB} . The frequency that each pair of mutations occurred together was also determined and termed P_{AB} , and the frequency that neither occurred was determined and termed P_{ab} . We first determined co-occurrence of two mutations by calculating standard linkage disequilibrium (D) by the formula: $D = (P_{AB} \times P_{ab}) - (P_{Ab} \times P_{aB})$. A normalized linkage disequilibrium (D') was calculated by dividing positive D values by a theoretical maximum, and negative D values by a theoretical minimum²¹. As our ultimate measure of correlation, r^2 (the square of the correlation coefficient) was calculated by the equation $r^2 = D^2 / (P_a P_b P_A P_B)$, where P_a and P_b are the frequency of all other bases at each of the nucleotide position of the pair of mutations under consideration. The value of r^2 multiplied by the number of sequences analysed is numerically equivalent to the value of χ^2 , which was used to determine statistical significance²¹. We note that molecules in our populations are not explicitly subject to recombination³⁵, such that one would not expect correlations between mutations to decay significantly in the relatively short time scales of our experiments.

Principal component analysis. Sequences from corresponding generations (1st, 4th and 8th) of lines New-A, New-B and New-WT were combined and aligned in a single multiple sequence alignment. Alignment sites containing more than 95% gaps were removed. Multiple sequence alignments were represented numerically (that is, gap:0; A:1; C:2; U:3; G:4), so a single sequence can be interpreted as a vector of n variables where each correspond to a nucleotide sites along the sequence. Principal component analysis was carried out using the princomp function in Matlab (The MathWorks).

Network graph. Sequences from generation 10 of line B, and the first generations of lines New-B and New-WT were combined and aligned as described above. The number of nucleotide differences between every pair of sequences in the alignment was counted (all against all distances). It was necessary to control for the number of nodes and edges to allow visualization of the full data set. Thus, sequences were clustered according to percentage identity using the cd-hit algorithm³⁶. The representative sequence for each cluster was defined as the sequence within that cluster with the lowest average distance to all other sequences in the cluster. Network graphs were constructed with clusters represented as nodes, and edges connecting clusters containing sequences that differ by less than 10 nucleotides. Networks representation was accomplished using Cytoscape v2.7.0 (www.cytoscape.org).

Phylogeny. Sequences that clustered (96% identity) with the Azo* sequences from generation B-10 and New-B1, as well as the representative sequences from generation New-WT1 (less than 96% identity with Azo*) were used to construct an unrooted phylogenetic tree using maximum likelihood, and assuming a HKY85 model, using the PhyML software³⁷. The sequence data was used to determine base frequencies and to estimate the ratio of transitions to transversions.

31. Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *Genome Res.* **2**, 28–33 (1992).
32. Eckert, K. A. & Kunkel, T. A. DNA polymerase fidelity and the polymerase chain reaction. *Genome Res.* **1**, 17–24 (1991).
33. Breaker, R. R. & Joyce, G. F. Inventing and improving ribozyme function: Rational design versus iterative selection methods. *Trends Biotechnol.* **12**, 268–275 (1994).
34. Needleman, S. B. & Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443–453 (1970).
35. Yu, W., Rusterholtz, K. J., Krummel, A. T. & Lehman, N. Detection of high levels of recombination generated during PCR amplification of RNA templates. *Biotechniques* **40**, 499–507 (2006).
36. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
37. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).

Non-apoptotic role of BID in inflammation and innate immunity

Garabet Yeretssian¹, Ricardo G. Correa², Karine Doiron¹, Patrick Fitzgerald³, Christopher P. Dillon³, Douglas R. Green³, John C. Reed² & Maya Saleh¹

Innate immunity is a fundamental defence response that depends on evolutionarily conserved pattern recognition receptors for sensing infections or danger signals^{1,2}. Nucleotide-binding and oligomerization domain (NOD) proteins are cytosolic pattern-recognition receptors of paramount importance in the intestine, and their dysregulation is associated with inflammatory bowel disease^{3,4}. They sense peptidoglycans from commensal microorganisms and pathogens and coordinate signalling events that culminate in the induction of inflammation and anti-microbial responses². However, the signalling mechanisms involved in this process are not fully understood. Here, using genome-wide RNA interference, we identify candidate genes that modulate the NOD1 inflammatory response in intestinal epithelial cells. Our results reveal a significant crosstalk between innate immunity and apoptosis and identify BID, a BCL2 family protein, as a critical component of the inflammatory response. Colonocytes depleted of BID or macrophages from *Bid*^{-/-} mice are markedly defective in cytokine production in response to NOD activation. Furthermore, *Bid*^{-/-} mice are unresponsive to local or systemic exposure to NOD agonists or their protective effect in experimental colitis. Mechanistically, BID interacts with NOD1, NOD2 and the I κ B kinase (IKK) complex, impacting NF- κ B and extracellular signal-regulated kinase (ERK) signalling. Our results define a novel role of BID in inflammation and immunity independent of its apoptotic function, furthering the mounting evidence of evolutionary conservation between the mechanisms of apoptosis and immunity.

To identify novel regulators of the NOD1 pathway, we conducted a small interfering RNA (siRNA) screen using a library targeting 7,170 human druggable genes. The human intestinal epithelial cell-line HT29 was used and interleukin (IL)-8 was measured as a readout⁵⁻⁷. A schematic of the screening strategy is presented in Fig. 1a. Our primary screen identified 227 and 198 genes that respectively enhanced or dampened IL-8 protein levels in response to the NOD1 agonist γ Tri-DAP (Fig. 1b and Supplementary Tables 1 and 2). To reduce potential off-target hits, we performed a validation screen by measuring IL-8 messenger RNA levels. A strong correlation between the primary and validation screens was observed (Fig. 1c), with 200 candidates retained independently of cell toxicity (Fig. 1c, d, Supplementary Fig. 1a, b and Supplementary Table 3). Next, we performed a counter-screen using tumour-necrosis factor (TNF) as an inducer of IL-8, because NOD1/2 and tumour necrosis factor receptor (TNFR1) pathways are structurally and functionally related^{8,9}. Altogether, 114 genes were identified as common candidate regulators of the NOD1/2 and TNFR1 pathways (Supplementary Fig. 1c and Supplementary Table 4) of which 60 were common to the NOD1 validated hits (Fig. 1d, Supplementary Fig. 1d and Supplementary Table 5).

To gain insights into the biological processes regulated by the uncovered candidates, we used PANTHER to cluster the primary hits according to their gene ontology (Supplementary Fig. 1e, f and Supplementary Tables 6–9). Genes involved in signal transduction,

metabolism and proliferation were significantly over-represented. Interestingly, genes of the apoptosis machinery were also significantly enriched. Analysis of the Novartis GNF human expression atlas data across 80 tissues showed that the candidate genes are predominantly expressed in immune and neuronal tissues (Wilcoxon $P < 0.05$; Supplementary Fig. 2a and Supplementary Tables 10–12). Furthermore, examination of their promoter region revealed significant enrichment in consensus sites for pro-inflammatory and mitogenic transcription factors (Supplementary Fig. 2b). These results are consistent with the established role of the NOD1 pathway in inflammation and IEC proliferation and support the validity of our screen. To define networks of interacting proteins, we used Ingenuity Pathway Analysis and manual data mining to construct an interaction network by anchoring on NOD1 pathway components. Network modelling identified functional groups linked to signal transduction, including known effectors of NOD1 signalling. Further analysis of this network identified the BH3-only BCL2 family member BID as a potential proximal regulator of the NOD1 signalling complex (Supplementary Fig. 2c).

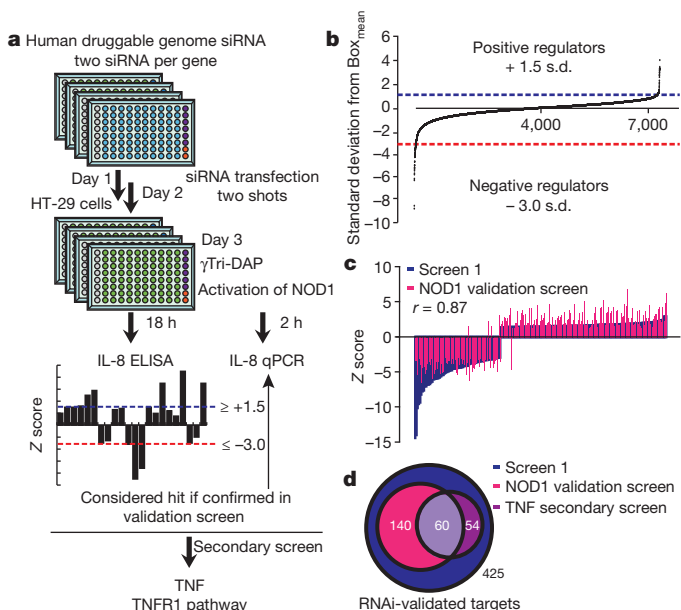


Figure 1 | Genome-wide RNA interference screen for genes regulating NOD1 signalling. **a**, Representation of the RNA interference screen strategy. **b**, Distribution of the data from the NOD1 primary screen. Each dot represents IL-8 concentration after individual gene depletion by a single siRNA pool. Data are represented as Z scores, with cut-offs (dashed lines) of 1.5 s.d. above the mean and 3.0 s.d. below the mean of all plates. **c**, Ranking by Z scores of 205 siRNAs from the primary and validation screens with a correlation coefficient $r = 0.87$. **d**, Venn diagram of hits from the primary, validation and secondary screens.

¹Department of Medicine, McGill University, Montreal, Quebec H3G 0B1, Canada. ²Sanford-Burnham Medical Research Institute, La Jolla, California 92037, USA. ³St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA.

Mounting evidence points to a co-evolution of the apoptosis and innate immunity pathways^{10–13}. Notably, NOD proteins and NOD-like receptors are structurally related to apoptosis-activating factor 1 (APAF1)^{10,11}, and the anti-apoptotic proteins BCL2 and BCL-X_L down-regulate innate immunity signalling by binding directly to NOD-like receptor proteins¹³. The interaction of BCL2 proteins with NOD-like receptors is reminiscent of that of CED9 with CED4, their respective orthologues in the *Caenorhabditis elegans* apoptosis pathway¹⁴.

To determine the role of BID in NOD1 signalling, we first validated the effect of BID depletion on γ Tri-DAP-induced IL-8 production in HT29 cells. Depletion of BID with three independent siRNAs led to a marked reduction in IL-8 production (Fig. 2a) similar to siRNAs against known positive regulators of the pathway (Fig. 2b and Supplementary Fig. 3a). Similarly, TNF and MCP-1 levels were dampened in BID-depleted cells in response to NOD1 activation (Supplementary Fig. 3a). Notably, depletion of other BCL2 family members including BCL2 or BAX did not modulate NOD signalling (Supplementary Fig. 3b). Consistent with the secondary screen results, BID was also required for TNF-induced IL-8 production (Supplementary Fig. 3c).

We next investigated the effect of BID depletion on the NOD1 signalling pathways. Figure 2c shows that BID impacted NF- κ B and ERK signalling. However, it was dispensable for p38 and Jun amino (N)-terminal kinases (JNK) activation (Supplementary Fig. 3d). The role of full-length BID is unclear and its function in apoptosis is contingent on its processing¹⁵. Furthermore, BID phosphorylation by

casein kinases I and II on serine residues proximal to the cleavage site hampers its processing and promotes cell survival¹⁶. Interestingly, γ Tri-DAP treatment of HT29 cells did not induce BID cleavage but resulted in its phosphorylation (Fig. 2c).

NOD1 is expressed in epithelial cells and fibroblasts, whereas NOD2 is highly expressed in myeloid cells¹⁷. We thus stimulated primary bone marrow-derived macrophages (BMDMs) from wild-type, *Ripk2*^{-/-} or *Bid*^{-/-} mice with the NOD2-specific agonist muramyl-dipeptide (MDP) as well as various Toll-like receptor (TLR) ligands and examined NOD2 signalling. IL-6 production was significantly blunted in BID-deficient macrophages in response to MDP alone or in synergy with lipopolysaccharide (LPS), but not TLR ligands (Fig. 2d and Supplementary Fig. 3e, f). Similarly, MDP-stimulated production of TNF was dampened in *Bid*^{-/-} BMDMs compared with wild-type cells (Supplementary Fig. 3e). As in BID-depleted HT29 cells, *Bid*^{-/-} macrophages failed to activate NF- κ B and ERK signalling in response to NOD stimulation (Fig. 2e), whereas JNK and p38 activation was normal (Supplementary Fig. 3d). The role of BID in signal transduction in intestinal epithelial cells and macrophages was independent of apoptosis induction by NOD1 or NOD2 agonists (Supplementary Fig. 3g). These results are consistent with recent findings that identified a role of BID in the activation of the NF- κ B pathway and cell survival¹⁸.

To explore the role of BID processing and its BH3 domain in inflammation, we reconstituted *Bid*^{-/-} BMDMs with either wild-type BID, a non-cleavable form of BID (BID_{D59E}) or an apoptosis-deficient

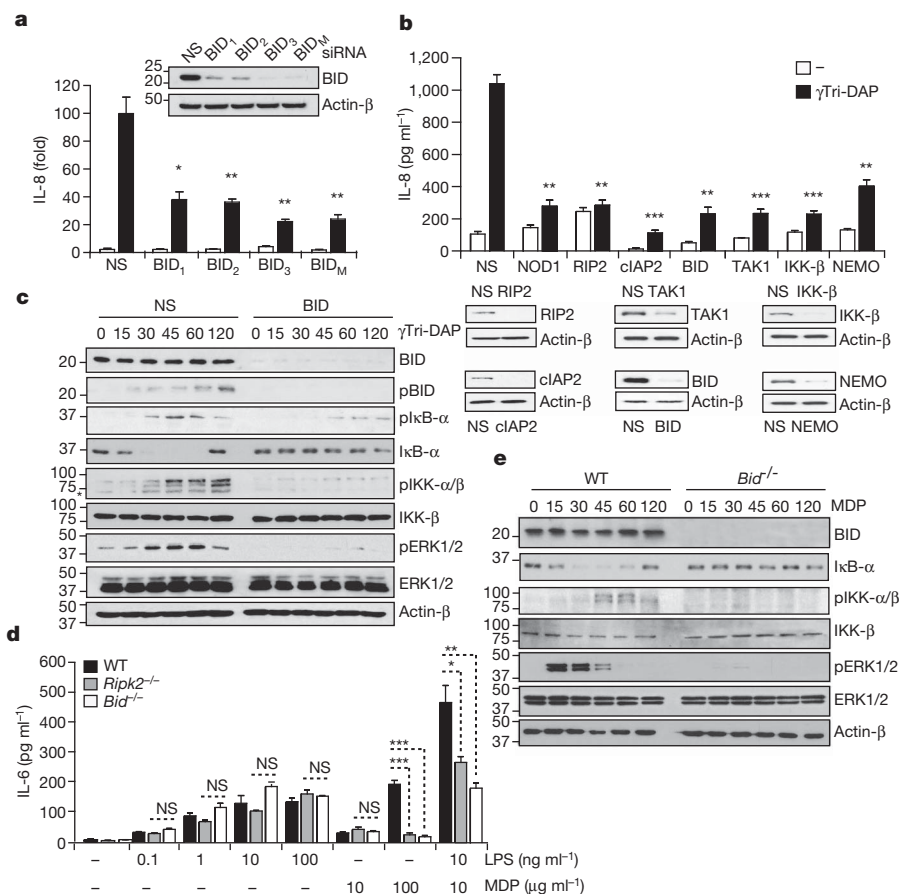


Figure 2 | BID is required for NOD1 and NOD2 signalling. **a**, HT29 cells were transfected with non-specific (NS) or three BID siRNA (BID₁, BID₂ or BID₃) alone or mixed (BID_M). IL-8 levels were determined by qPCR 2 h after DAP treatment. **b**, HT29 cells were treated as in **a**. IL-8 production was quantified by ELISA 18 h after DAP treatment. Protein depletion was assessed by western blotting (insets in **a** and **b**). **c**, HT29 cells were treated as in **a** for the indicated time points. I κ B α degradation and phosphorylation and MAPK, IKK α / β and BID phosphorylation were assayed by immunoblotting.

d, BMDMs from wild-type (WT, $n = 4$), *Ripk2*^{-/-} ($n = 4$) and *Bid*^{-/-} ($n = 4$) mice were left untreated or stimulated with LPS, MDP or MDP with LPS for 18 h. IL-6 production was quantified by ELISA. **e**, BMDMs from wild-type and *Bid*^{-/-} mice were treated with MDP. I κ B α degradation and MAPK, IKK α / β and BID phosphorylation were assayed by immunoblotting. p, phosphorylated. NS, not significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. Values in graphs represent mean \pm s.e.m. from three or four independent experiments.

mutant form of BID (BID_{G94E})¹⁹, and examined cytokine production induced by MDP. Figure 3a shows that BID, BID_{D59E} and BID_{G94E} were equivalent in rescuing the blunted inflammatory phenotype of $Bid^{-/-}$ cells. These results suggested that the role of BID in inflammation is independent of its processing, its BH3 domain or its association with BAX. This was further confirmed in $Bax^{-/-}Bak^{-/-}$ macrophages from tamoxifen-treated $ROSA-CreER Bax^{fl/-}Bak^{fl/-}$ mice that showed an unaltered response to MDP compared with wild-type cells (Fig. 3b). In contrast, mutation of reported phosphorylation sites in BID (BID_{S64A} , BID_{S65A} and BID_{S78A}) significantly blunted MDP-induced IL-6 production (Fig. 3a).

To examine whether BID associated with components of the NOD signalosome, we performed co-immunoprecipitation experiments and showed that BID interacted with NOD1 and NOD2 (Supplementary Fig. 4a). Similar results were obtained with BID_{G94E} (Supplementary Fig. 4b), but not with the serine to alanine mutants of BID, which were deficient in NOD1 binding (Supplementary Fig. 4c). *In vitro* pull-down experiments indicated that binding of NOD1 and NOD2 to purified recombinant BID was mediated through the central region encompassing the nucleotide-binding domain (Supplementary Fig. 4d). This result

is comparable to the binding of BCL2/BCL-X_L to NLRP1 (ref. 13). The region in BCL2/BCL-X_L required for this association has been previously mapped to an unstructured loop between α -helices 1 and 2 (ref. 20). Interestingly, full-length BID contains such a loop (residues 40–80)²¹. BID did not interact with NLRP1 (Supplementary Fig. 4e); however, truncation of its loop markedly reduced its association with NOD1 and NOD2 (Supplementary Fig. 4f). Consistently, subtle perturbation of the helical structure around the loop, through mutagenesis of specific leucine residues (L37A, L40A and L44A), drastically weakened the BID–NOD1 interaction (Supplementary Fig. 4g).

Upon agonist sensing, NOD1 and NOD2 associate with RIP2 and TAK1, which leads to activation of NF- κ B and MAPK (ERK1/2, p38 and JNK) pathways². The bifurcation in signalling between the NF- κ B and ERK versus p38 and JNK pathways occurs downstream of TAK1 and might involve differential recruitment of adaptor proteins that direct signalling towards specific inflammatory responses. TAK1 binds to IKK- γ (NEMO), MKK3/6 and MKK4/7 to trigger NF- κ B, p38 and JNK activation, respectively; whereas signalling to ERK1/2 occurs through an interaction between IKK- β and tumour progression locus 2 (TPL2)^{22–24}. Notably, the adaptor protein CARD9 transduces NOD signalling to p38 and JNK but not the NF- κ B pathway²⁵. The distinct requirement of BID in NF- κ B and ERK activation points to a possible interaction of BID with a signalling protein that associates with the IKK complex. Co-immunoprecipitation experiments revealed that BID associates with protein complexes containing IKK α , IKK- β and NEMO (Supplementary Fig. 4h). Reciprocally, NEMO associated with protein complexes containing BID and RIP2 (Supplementary Fig. 4i). To support these results further, we examined the association of endogenous BID with the NOD complex. Figure 3c shows that endogenous BID interacted with RIP2 in an agonist-dependent manner. Additionally, recruitment of the IKK complex to the NOD signalosome was mediated by BID, as NEMO interacted with RIP2 in wild-type but not $Ripk2^{-/-}$ or $Bid^{-/-}$ BMDMs after MDP stimulation (Fig. 3c). Immunoprecipitation of endogenous NEMO also revealed interaction with BID, IKK- β and RIP2-containing complexes in response to NOD activation (Fig. 3c). Altogether, these biochemical findings suggest that BID bridges the NOD receptors to the IKK complex (Fig. 3d).

To examine the function of BID in inflammation and innate immunity *in vivo*, we assessed the systemic and local responses of wild-type, $Ripk2^{-/-}$ and $Bid^{-/-}$ mice to administration of NOD agonists. Consistent with previous reports^{26,27}, intraperitoneal challenge with γ Tri-DAP induced a robust systemic production of chemokines and cytokines (Fig. 4a) and a pronounced peritonitis marked by infiltration of Gr1⁺ neutrophils (Fig. 4b) in wild-type mice but not $Ripk2^{-/-}$ or $Bid^{-/-}$ mice (Fig. 4a, b). Similarly, IL-6 production was dampened in $Ripk2^{-/-}$ and $Bid^{-/-}$ mice compared with wild-type mice in response to MDP administration (Supplementary Fig. 5a) but not that of LPS (Supplementary Fig. 5b). It has been previously shown that administration of the TLR3 ligand poly I:C and the NOD2 agonist MDP protect mice from dextran-sulphate sodium (DSS)-induced colitis^{26,28}. The protective effect of MDP is abrogated in mice deficient in NOD2, RIP2 or cIAP2 (refs 26, 27). To explore the role of BID in the NOD2 pathway, we subjected wild-type, $Ripk2^{-/-}$ and $Bid^{-/-}$ mice to an acute DSS treatment. Apart from wild-type mice treated with MDP, all other mice showed severe weight loss (Fig. 4c) and thickening and shortening of the colons upon necropsy (Supplementary Fig. 5d, e). Histological examination of colon sections further supported these data, as wild-type mice treated with MDP and DSS showed evidence of restored colonic crypt architecture, whereas all other mice exhibited severe leukocyte recruitment and crypt loss (Supplementary Fig. 5f). In contrast, $Ripk2^{-/-}$ and $Bid^{-/-}$ mice reacted similarly to wild-type mice in response to poly I:C (Supplementary Fig. 5c–f). Thus loss of BID function impaired the ability of NOD2 but not TLR3 to trigger tissue repair responses upon injury with DSS.

In summary, our high-throughput functional screen has identified approximately 200 candidate modulators of the NOD1 pathway and

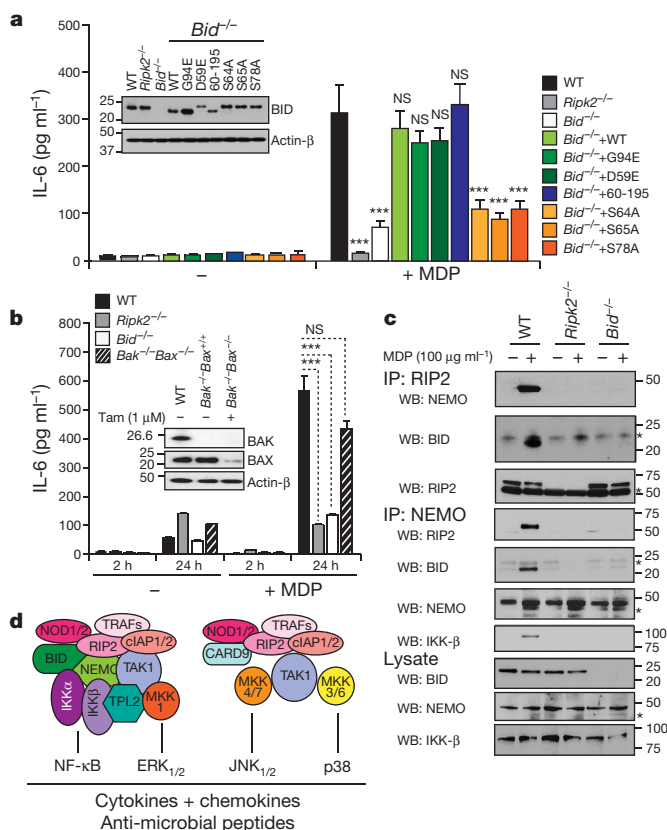


Figure 3 | BID regulates NOD signalling independently of its apoptotic determinants by linking NOD proteins to the IKK complex. **a**, BMDMs from $Bid^{-/-}$ mice ($n = 4$) were reconstituted with control or full-length BID, BID_{G94E} , BID_{D59E} , BID_{60-195} , BID_{S64A} , BID_{S65A} or BID_{S78A} mutants and stimulated with MDP for 18 h. IL-6 production was quantified by ELISA. BID expression was determined by immunoblotting (inset). **b**, BMDMs were isolated from wild-type ($n = 4$), $Ripk2^{-/-}$ ($n = 4$), $Bid^{-/-}$ ($n = 4$) and $ROSA-CreER Bax^{fl/-}Bak^{fl/-}$ ($n = 3$) mice. 4-OH tamoxifen was used to excise the *Bax* gene. BMDMs were challenged with MDP for the indicated time points and IL-6 production was measured by ELISA. The efficiency of *Bax* excision was determined by immunoblot (inset). **c**, BMDMs from wild-type, $Ripk2^{-/-}$ and $Bid^{-/-}$ mice were left untreated or treated with MDP for 30 min. Cell lysates were used for immunoprecipitation using anti-RIP2 or NEMO antibodies and analysed by immunoblotting. NS, not significant; *** $P < 0.001$. Asterisk, IgG heavy chain. Values in graphs represent mean \pm s.e.m. from three independent experiments. **d**, Representation of the NOD signalosome.

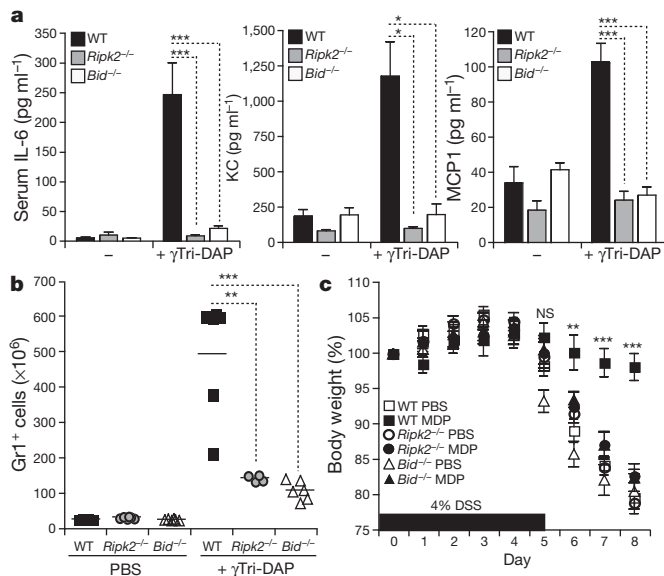


Figure 4 | *Bid*^{-/-} mice exhibit a blunted inflammatory response after NOD stimulation *in vivo* and are not protected from DSS colitis by MDP. **a**, Wild-type (WT, *n* = 6), *Ripk2*^{-/-} (*n* = 4) and *Bid*^{-/-} (*n* = 6) mice were injected intraperitoneally with γTri-DAP or endotoxin-free PBS for 2 h. IL-6, KC and MCP-1 levels in the serum were measured by ELISA. **b**, Mice were injected intraperitoneally with γTri-DAP for 24 h. The number of infiltrating Gr1⁺ cells in the peritoneum was quantified by fluorescence-activated cell sorting. **c**, Wild-type (*n* = 8), *Ripk2*^{-/-} (*n* = 8) and *Bid*^{-/-} (*n* = 8) mice were treated with 4% DSS. On days 0, 1 and 2, mice were injected intraperitoneally with MDP or endotoxin-free PBS. Body weight loss is represented. NS, not significant; **P* < 0.05; ***P* < 0.01; ****P* < 0.001. Values in graphs represent mean ± s.e.m. from three independent experiments.

provides a comprehensive view of the biological processes involved in NOD signalling. Importantly, we have uncovered the physiological function of full-length BID as a critical mediator of inflammation and innate immunity. Our results place BID at the centre of the cell's decision to induce innate immune responses (and survive) or commit suicide by apoptosis. We anticipate that the development of small molecules targeting BID will set the stage for the emergence of novel strategies to treat immune-mediated inflammatory diseases including inflammatory bowel disease.

METHODS SUMMARY

The siRNA screen was performed using the human druggable genome siRNA set V2.0 library from Qiagen. HT29 cells were transfected with siRNA (50 nM) using Lipofectamine 2000 and treated with DOTAP conjugated γTri-DAP (10 μg ml⁻¹). Cytokines and chemokines were measured by enzyme-linked immunosorbent assay (ELISA) and quantitative real-time PCR (qPCR). Data were analysed using *cellHTS2*. Bioinformatics were analysed using PANTHER, MSigDB and TRANSFAC, and network modelling was performed with data obtained from the Human Protein Reference Database, the Biomolecular Interaction Network Database and the Ingenuity Pathways Analysis database. BMDMs were treated with DOTAP-conjugated MDP (10–100 μg ml⁻¹). Reconstitution of BMDMs was performed with Amara (Solution T, program T-20). NOD signalling was assessed by immunoblotting using total and phospho-specific antibodies. Co-immunoprecipitation experiments were performed in HEK293T cells. γTri-DAP (1 mg kg⁻¹), MDP (100 μg) and poly I:C (100 μg) were injected intraperitoneally in mice, and 4% DSS was administered in the drinking water.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 August 2010; accepted 3 March 2011.

Published online 8 May 2011.

- Garrett, W. S., Gordon, J. I. & Glimcher, L. H. Homeostasis and inflammation in the intestine. *Cell* **140**, 859–870 (2010).

- Takeuchi, O. & Akira, S. Pattern recognition receptors and inflammation. *Cell* **140**, 805–820 (2010).
- Cho, J. H. The genetics and immunopathogenesis of inflammatory bowel disease. *Nature Rev. Immunol.* **8**, 458–466 (2008).
- Turner, J. R. Intestinal mucosal barrier function in health and disease. *Nature Rev. Immunol.* **9**, 799–809 (2009).
- Girardin, S. E. *et al.* Nod1 detects a unique muropeptide from Gram-negative bacterial peptidoglycan. *Science* **300**, 1584–1587 (2003).
- Kim, J. G., Lee, S. J. & Kagnoff, M. F. Nod1 is an essential signal transducer in intestinal epithelial cells infected with bacteria that avoid recognition by toll-like receptors. *Infect. Immun.* **72**, 1487–1495 (2004).
- Masumoto, J. *et al.* Nod1 acts as an intracellular receptor to stimulate chemokine production and neutrophil recruitment *in vivo*. *J. Exp. Med.* **203**, 203–213 (2006).
- Chen, C. M., Gong, Y., Zhang, M. & Chen, J. J. Reciprocal cross-talk between Nod2 and TAK1 signaling pathways. *J. Biol. Chem.* **279**, 25876–25882 (2004).
- Kobayashi, K. S. *et al.* Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* **307**, 731–734 (2005).
- Bertin, J. *et al.* Human CARD4 protein is a novel CED-4/Apaf-1 cell death family member that activates NF-κB. *J. Biol. Chem.* **274**, 12955–12958 (1999).
- Inohara, N. *et al.* Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor-κB. *J. Biol. Chem.* **274**, 14560–14567 (1999).
- Faustin, B. *et al.* Reconstituted NALP1 inflammasome reveals two-step mechanism of caspase-1 activation. *Mol. Cell* **25**, 713–724 (2007).
- Bruey, J. M. *et al.* Bcl-2 and Bcl-XL regulate proinflammatory caspase-1 activation by interaction with NALP1. *Cell* **129**, 45–56 (2007).
- Conradt, B. & Horvitz, H. R. The *C. elegans* protein EGL-1 is required for programmed cell death and interacts with the Bcl-2-like protein CED-9. *Cell* **93**, 519–529 (1998).
- Strasser, A. The role of BH3-only proteins in the immune system. *Nature Rev. Immunol.* **5**, 189–200 (2005).
- Desagher, S. *et al.* Phosphorylation of bid by casein kinases I and II regulates its cleavage by caspase 8. *Mol. Cell* **8**, 601–611 (2001).
- Xavier, R. J. & Podolsky, D. K. Unravelling the pathogenesis of inflammatory bowel disease. *Nature* **448**, 427–434 (2007).
- Luo, W. *et al.* Bid mediates anti-apoptotic COX-2 induction through the IKKβ/NFκB pathway due to 5-MCDE exposure. *Curr. Cancer Drug Targets* **10**, 96–106 (2010).
- Wang, K., Yin, X. M., Chao, D. T., Milliman, C. L. & Korsmeyer, S. J. BID: a novel BH3 domain-only death agonist. *Genes Dev.* **10**, 2859–2869 (1996).
- Faustin, B. *et al.* Mechanism of Bcl-2 and Bcl-X(L) inhibition of NLRP1 inflammasome: loop domain-dependent suppression of ATP binding and oligomerization. *Proc. Natl Acad. Sci. USA* **106**, 3935–3940 (2009).
- Petros, A. M., Olejniczak, E. T. & Fesik, S. W. Structural biology of the Bcl-2 family of proteins. *Biochim. Biophys. Acta* **1644**, 83–94 (2004).
- Hasegawa, M. *et al.* A critical role of RICK/RIP2 polyubiquitination in Nod-induced NF-κB activation. *EMBO J.* **27**, 373–383 (2007).
- Abbott, D. W., Wilkins, A., Asara, J. M. & Cantley, L. C. The Crohn's disease protein, NOD2, requires RIP2 in order to induce ubiquitylation of a novel site on NEMO. *Curr. Biol.* **14**, 2217–2227 (2004).
- Perkins, N. D. Integrating cell-signalling pathways with NF-κB and IKK function. *Nature Rev. Mol. Cell Biol.* **8**, 49–62 (2007).
- Hsu, Y. M. *et al.* The adaptor protein CARD9 is required for innate immune responses to intracellular pathogens. *Nature Immunol.* **8**, 198–205 (2007).
- Watanabe, T. *et al.* Muramyl dipeptide activation of nucleotide-binding oligomerization domain 2 protects mice from experimental colitis. *J. Clin. Invest.* **118**, 545–559 (2008).
- Bertrand, M. J. *et al.* Cellular inhibitors of apoptosis cIAP1 and cIAP2 are required for innate immunity signaling by the pattern recognition receptors NOD1 and NOD2. *Immunity* **30**, 789–801 (2009).
- Vijay-Kumar, M. *et al.* Activation of toll-like receptor 3 protects against DSS-induced acute colitis. *Inflamm. Bowel Dis.* **13**, 856–864 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Strasser for providing *Bid*^{-/-} mice, G. Shore for BID antibodies and the McGill University high throughput/high content screening facility. We also thank D. Zhai for BID purification. This work was supported by grants from the Canadian Institutes for Health Research (CIHR-MOP 82801) and the Burroughs Wellcome Fund to M.S. M.S. is a Canadian Institutes for Health Research New Investigator. G.Y. is supported by a PDF-Fellowship from the McGill University Health Center. C.P.D. is supported by a fellowship grant from the SASS Foundation for Medical Research.

Author Contributions G.Y. and M.S. designed the research. G.Y., K.D. and M.S. performed the screen. G.Y. and M.S. analysed the data. G.Y. performed most experiments. R.G.C. performed Ni/NTA pull-down assay; P.F., R.G.C., C.P.D., D.R.G. and J.C.R. contributed new reagents/analytical tools. G.Y. and M.S. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.S. (maya.saleh@mcgill.ca).

METHODS

Animal strains. Wild-type (C57Bl/6) *Bid*^{-/-} mice from A. Strasser and *Ripk2*^{-/-} from R. Flavell were backcrossed to a pure C57Bl/6 genetic background. Housing and all animal procedures were approved by McGill University operating under the guidelines of the Canadian Council of Animal Care.

Reagents and plasmids. Antibodies for pJNK (9255), pp38 (9211), p38 (9212), pERK1/2 (9106S), ERK1/2 (9102), IκBα (4812), pIκBα (9246 s), IKK-β (2370) and pIKK-α/β (2697) were purchased from Cell Signaling Technologies. Anti-actin-β (A1978) was from Sigma; anti-JNK1/3 (sc-474), anti-RIP2 (sc-22763), anti-BID (sc-11423), anti-BAX (sc-493), anti-NEMO (sc-8330) and anti-BAK (sc-832) were from Santa Cruz Biotechnology. Monoclonal anti-RIP2 was obtained from BD Bioscience (612349), anti-TAK1 (7263) from Millipore and cIAP2-specific antibody from R&D systems (AF8181). Antibody against pBid-S61 (A300-527A) was obtained from Bethyl laboratories. The Flag antibody was purchased from Sigma (F3165). All other antibodies against tags anti-haemagglutinin (11583816001), anti-Myc (11667149001) and anti-green fluorescent protein (11814460001) were obtained from Roche. Secondary antibodies were from Jackson ImmunoResearch Laboratories. All siRNAs used in this study were purchased from Qiagen. γTri-DAP (60774) was from Anaspec, MDP (G-1055) was from Bachem, and human recombinant TNF-α (30001A) and mouse recombinant TNF-α (31501) were from Peprotech. Cyclohexamide (C4859) was from Sigma and AnnexinV-FITC (556420) was from BD Pharmingen. DOTAP transfection reagent (11202375001) was from Roche, and DSS (relative molecular mass 36,000–50,000), was from MP Biomedicals (160110).

To generate the non-cleavable form of BID (*BID*_{D59E}), the apoptosis-deficient mutant form of BID (*BID*_{G94E}), the serine to alanine mutant forms of BID (*BID*_{S64A}, *BID*_{S65A} and *BID*_{S78A}) as well as the leucine to alanine mutant forms of BID (*BID*_{L37A}, *BID*_{L40A} and *BID*_{L44A}) mutagenesis PCRs on pcDNA3.1-HA-BID or pEGFP-BID were performed using a site-directed mutagenesis kit from Stratagene (200519-5). HA-tagged BID deletion constructs were obtained by PCR sub-cloning into pcDNA3.1 with the primers listed in Supplementary Table 13. The plasmids expressing Flag-tagged IKK-α and IKK-β were obtained from Tularik. The Flag-tagged NOD2 construct was obtained by inserting a Flag sequence in frame with the human NOD2 sequence in a pcDNA3.1 plasmid originally from G. Nunez. Myc-tagged NLRP1 was described previously^{9,16}. Myc-tagged NEMO was obtained from I. Verma. The construct expressing Flag-tagged RIP2 was a gift of G. Nunez; Flag-NOD1, RIP2-CARD and RIP2-ΔCARD plasmids were described previously²⁵. The plasmid encoding HA-TAK1 was provided by S. Baksh. Myc-tagged NOD1 and NOD2, as well as their deletion constructs, were described previously²⁹.

Cell culture. HEK293T and HT29 cells were maintained in 5% CO₂ at 37 °C in Dulbecco's modified Eagle's medium (HyClone) and McCoy's 5A modified medium (HyClone), respectively, supplemented with 10% fetal calf serum (HyClone), 2 mM L-glutamine and 100 μg ml⁻¹ penicillin/streptomycin. siRNA as well as plasmids were transfected with Lipofectamine 2000 (Invitrogen, 11668-19). We treated HT29 cells for different time points (0–120 min) with γTri-DAP at 10 μg ml⁻¹ transfected into the cells by using the DOTAP cationic lipid transfection reagent in accordance with the manufacturer's instructions. Cell-culture supernatants were used to quantify IL-8 levels by ELISA. For cell-death assay, HT29 cells were treated with γTri-DAP (10 μg ml⁻¹) or TNF-α (10 ng ml⁻¹) alone or in combination with cyclohexamide (10 μg ml⁻¹) for 24 h and 48 h.

siRNA screen and data analysis. siRNA screen was performed using the human druggable genome siRNA set V2.0 library from Qiagen and the Biomek II liquid handler from Beckman Coulter. The library consists of a pool of two siRNA duplexes for each target gene. A total of 7,170 human genes were assayed in the primary screen. For the validation screen, siRNAs for each gene used in the first screen were procured separately and tested. The library was received in 89-well × 96-well plates at 5 μM stock concentration. We prepared daughter plates at 1 μM that we used in the screen. HT29 cells (5 × 10³ per well) were transfected with siRNA (50 nM) using Lipofectamine 2000. After 24 h a second siRNA transfection was performed to ensure high efficiency of knockdown, and plates were incubated for an additional 24 h before treatment with DOTAP-encapsulated γTri-DAP (10 μg ml⁻¹) or recombinant human TNF-α (10 ng ml⁻¹). For the primary and secondary screens, cells were treated for 16 h with γTri-DAP or TNF-α and culture supernatants collected to measure IL-8 production by ELISA. For the validation screen, cells were treated for 2 h with γTri-DAP and IL-8 messenger RNA levels were quantified by qPCR. The raw data obtained from the primary screen were analysed using the open source Bioconductor/R package, *cellHTS2* (<http://www.dkfz.de/signaling/cellHTS>)³⁰. This program generated annotated gene lists with the corresponding adjusted Z scores (s.d. from the mean value of the control wells). We defined hits as positive or negative regulators according to a Z-score cut-off. Our cut-off criterion was 1.5 s.d. above the mean value for positive regulators and 3.0 s.d. below the mean for negative regulators. To evaluate the robustness and sensitivity of our screen, we calculated the overall Z'

factor of the screen as previously described³¹ taking into account the repeated positive and negative controls. The Z' factor obtained was greater than 0.5, confirming the overall quality of our assay.

Enrichment analyses. Genes were classified into biological process and molecular function categories using the PANTHER classification system³² and transcription-factor-binding sites using MSigDB and TRANSFAC v7.4 (www.gene-regulation.com). To assess the statistical enrichment or overrepresentation of these categories for the hit genes relative to their representation in the global set of genes examined in the siRNA screen, P values were computed using the hypergeometric test³³, which was implemented in the R language (<http://www.r-project.org/>). Briefly, the hypergeometric distribution describes the probability of finding at least *s* genes associated with a particular category in a set of *g* genes involved in NOD1 innate immune response, given that there are *S* genes associated with that same category in the global set of *G* genes examined in the genome-wide siRNA screen. For each category, *c*, and the list of genes *l*, the P value was calculated as $P(c, l) = 1 - \sum_{k \in \{0, 1, \dots, s\}} [C(g, k)C(G - g, S - k)/C(G, S)]$. The binomial coefficient is of the form $C(n, r)$. A value of *P* < 0.05 was considered significant. Categories assigned with at least ten genes are displayed in Supplementary Figs 2 and 3b.

Human protein interaction networks. The protein network was constructed by iteratively connecting interacting proteins, with data obtained from the Human Protein Reference Database, the Biomolecular Interaction Network Database, the Ingenuity Pathways database and functional information from the literature. The network uses graph theoretical representation in which components (gene products) are depicted as nodes and interactions between components as edges, implemented in the Perl programming language.

Analysis of gene expression across 80 tissues. Microarray data files were obtained from the Novartis GNF human expression atlas version-2 resource. Expression values of 33,689 probe sets from the HG-U133A (Affymetrix) platform and the GNF1H custom chip were analysed. The data set was normalized using global median scaling, and we filtered the data by excluding from the analysis those probe sets with 100% 'absent' calls (MAS5.0 algorithm) across all 80 tissues. The data set was further filtered by setting a minimum threshold value greater than 20 in at least one sample for each probe set and a maximum-mean expression value greater than 100. Hierarchical clustering (centroid linkage method) was performed with Cluster 3.0 using Pearson's correlation as the similarity metric³⁴. Z-score transformation was applied to each probe set across all arrays before generating 'heat maps' for visualization using TreeView³⁵.

Primary culture, transfection and treatment of bone-marrow-derived macrophages. Femurs and tibias of wild-type (C57Bl/6), *Ripk2*^{-/-}, *Bid*^{-/-} and *ROSA-CreER Bax*^{fl/fl} *Bak*^{-/-} mice were flushed with a 25G needle and cold RPMI 1640 (GIBCO) into a 15 ml tube. The suspension was then filtered through a 70 μm cell strainer and spun for 5 min at 400g. Pellets were re-suspended in complete RPMI 1640 supplemented with 15% L929 cell-conditioned media and plated in non-cell-culture-treated 10 cm dishes. Cells were cultured at 37 °C in a 10% CO₂ tissue culture incubator for 5–6 days, with media changed on day 3. MDP (at 10 or 100 μg ml⁻¹) was transfected into cells for different time points (0–120 min) with DOTAP reagent according to the manufacturer's instructions. BMDMs were transfected with BID wild-type and *BID*_{G94E} plasmids using the AMAXA Nucleofector (Solution T, program T-20; AMAXA biosystems) and treated with MDP (100 μg ml⁻¹). *ROSA-CreER Bax*^{fl/fl} *Bak*^{-/-} cells were subjected to 1 μM of 4-OH-tamoxifen (Sigma, T5648) during the differentiation period, followed by MDP treatment for 18 h. BMDMs were treated for 18 h with various TLR ligands: Pam3sk4 (1 μg ml⁻¹, TLR2), poly I:C (25 μg ml⁻¹, TLR3), LPS (*Escherichia coli* O111:B4; 0.1–100 ng ml⁻¹ and 1 μg ml⁻¹, TLR4), PGN (*Bacillus subtilis*; 2 μg ml⁻¹, flagellin (2 μg ml⁻¹, TLR5), resiquimod R848 (10 μg ml⁻¹, TLR7) and CpG (1 μg ml⁻¹, TLR9). In all these experiments, IL-6 production was measured in cell-culture supernatants using ELISA. For cell death assay, BMDMs were treated with MDP (100 μg ml⁻¹) or TNF-α (10 ng ml⁻¹) alone or in combination with cyclohexamide (10 μg ml⁻¹) for 24 h.

ELISA and lactate dehydrogenase assay. Cytokine and chemokine concentrations in cell-culture supernatants as well as sera were measured by ELISA using human IL-8/CXCL8 (DY208) and mouse IL-6 (DY406), KC/CXCL1 (DY453) and MCP-1/CC12 (DY479) kits purchased from R&D systems, used according to the manufacturer's instructions. For lactate dehydrogenase measurement, the CytoTox-96 non-radioactive cytotoxicity assay (Promega) was used according to the manufacturer's recommendations. The percentage of lactate dehydrogenase release was calculated compared with 100% cell lysis control.

Annexin V-FITC and propidium iodide staining. HT29 cells or BMDMs treated as described above were collected and re-suspended in 0.5 ml of Annexin V staining buffer (10 mM HEPES, 150 mM NaCl, 5 mM KCl, 2.5 mM CaCl₂, 100 μM MgCl₂). After adding 5 μl Annexin V-FITC, cells were incubated at room temperature for 15 min in the dark, propidium iodide (50 μg ml⁻¹) was added to each sample and cell death analysed by flow cytometry.

qPCR. Total RNA was extracted from cells with Trizol reagent (Invitrogen, 48190011), followed by isopropanol precipitation. We reverse-transcribed 2 µg total RNA to complementary DNA cDNA by using random hexamers and the M-MLV reverse transcriptase (Invitrogen, 28025-013) according to the manufacturer's protocol. qPCR with reverse transcription was performed using iTaq SYBR green supermix (Bio-Rad, 172-5852). The primers for different cytokines and chemokines used in this study are described in Supplementary Table 13.

Immunoblotting, immunoprecipitation and pull-down experiments. SDS-polyacrylamide gel electrophoresis and western blotting were performed in accordance with standard protocols. For signalling experiments, cells were lysed directly in Laemmli sample buffer. For co-immunoprecipitation experiments, HEK293T cells were co-transfected using Lipofectamine 2000. Cells were lysed in buffer B150 (20 mM Tris-HCl (pH 8.0), 150 mM KCl, 10% glycerol, 5 mM MgCl₂, and 0.1% NP-40) supplemented with a protease inhibitor tablet (Roche). Immunoassays were performed in RIPA buffer (10 mM Tris (pH 8.0), 150 mM NaCl, 1% Nonidet P-40, 0.1% SDS and 0.5% deoxycholate, supplemented with a protease inhibitor tablet (Roche)). In brief, whole cell lysates were collected 24 h later and incubated overnight with either M2 agarose beads (Sigma, A2220) or the immunoprecipitating antibody along with protein A/G sepharose beads (Sigma, P9424, P3296). Immunoprecipitates were eluted with Flag peptide (Sigma, F3290) or by boiling in Laemmli buffer. Eluted proteins were detected using monoclonal antibodies against green fluorescent protein, Flag, haemagglutinin or Myc tags. For direct pull-down assays, HEK293T cells were transfected with Myc-tagged NOD1, NOD2 or their deletion mutants. After 24 h, cells were lysed in WCE buffer (20 mM HEPES (pH 7.7), 500 mM NaCl, 1 mM EDTA, 0.25% Triton X-100, 1 mM EGTA) supplemented with 0.5 mM DTT and protease inhibitor cocktail (Roche). After extraction, protein samples were equilibrated in immunoprecipitation buffer (20 mM Tris (pH 8.0), 250 mM NaCl, 0.05% Nonidet P-40, 3 mM EDTA, 3 mM EGTA). Supernatants were collected and incubated for 16 h with purified recombinant histidine-tagged BID protein conjugated to Ni-NTA agarose beads (Qiagen). Immunoprecipitates were then analysed by western blotting. BID protein was produced in bacteria and purified by Ni-chelation chromatography similar to previous reports³⁶. For endogenous immunoprecipitation assays, 10⁸ BMDMs were used from wild-type, *Ripk2*^{-/-} or *Bid*^{-/-} mice. Cells were lysed in RIPA buffer supplemented with a protease inhibitor tablet (Roche), 5 mM sodium fluoride and 1 mM sodium orthovanadate. Lysates were suspended in 0.5 ml lysis solution, cleared by incubation with 15 µl protein A/G Sepharose beads, then incubated with 25 µl polyclonal antibody and 40 µl protein A/G beads at 4 °C overnight. Samples were then washed four times with lysis buffer, boiled in Laemmli buffer and analysed by SDS-polyacrylamide gel.

Intraperitoneal injection of LPS, MDP and γTri-DAP. Wild-type, RIP2-deficient and BID-deficient mice were injected intraperitoneally with 1 mg kg⁻¹ of γTri-DAP. Twenty-four hours after injection, infiltrating cells were collected in PBS by peritoneal lavage, counted and labelled for 30 min in the dark with FITC-labelled Ly6G/C antibody (BD Pharmingen, 553126). Gr-1⁺ cells were enumerated by flow cytometry using FACSCalibur (BD Biosciences). γTri-DAP (2 mg kg⁻¹) was injected intraperitoneally, blood was collected by heart puncture 2 h after agonist injection and serum used to quantify IL-6, KC and MCP-1 levels by ELISA. Similarly, MDP (100 µg; 500 µl per mouse) or LPS (10 mg kg⁻¹) were injected intraperitoneally. Blood was collected after 2 h and 4 h for MDP and 30 min for LPS. IL-6 levels were quantified by ELISA.

DSS colitis. Wild-type, *Ripk2*^{-/-} and *Bid*^{-/-} mice were given 4% DSS in drinking water for 6 days (days 0–5), then placed on regular water for 3 days (days 6–8). Mice were administered intraperitoneal MDP (100 µg in 500 µl PBS), poly I:C (100 µg in 500 µl PBS) or endotoxin-free PBS early during colitis on days 0, 1 and 2. Mice were monitored daily for body weight loss.

Histopathological analysis and immunohistochemistry. Colons were fixed in 10% formalin and embedded in paraffin. Serial sections (5 µm) were cut onto glass slides and stained with haematoxylin and eosin. Sections were blindly assessed for signs of colitis: tissue morphology, epithelial erosion, loss of goblet cells, loss of crypts, presence of inflammatory infiltrate and thickening of the colon wall.

Statistical analysis. A two-tailed Student's *t*-test was used to evaluate significant differences between two groups. One-way ANOVA was used to evaluate significant differences among multiple groups.

29. Ogura, Y. *et al.* Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-κB. *J. Biol. Chem.* **276**, 4812–4818 (2001).
30. Boutros, M., Bras, L. P. & Huber, W. Analysis of cell-based RNAi screens. *Genome Biol.* **7**, R66 (2006).
31. Zhang, J. H., Chung, T. D. & Oldenburg, K. R. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screen.* **4**, 67–73 (1999).
32. Mi, H. *et al.* The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* **33**, D284–D288 (2005).
33. Rivals, I. *et al.* Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics* **23**, 401–407 (2007).
34. Eisen, M. B. *et al.* Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
35. Saldanha, A. J. Java Treeview-extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248 (2004).
36. Zhai, D. *et al.* Humanin binds and nullifies Bid activity by blocking its activation of Bax and Bak. *J. Biol. Chem.* **280**, 15815–15824 (2005).

Glutamate induces *de novo* growth of functional spines in developing cortex

Hyung-Bae Kwon & Bernardo L. Sabatini

Mature cortical pyramidal neurons receive excitatory inputs onto small protrusions emanating from their dendrites called spines. Spines undergo activity-dependent remodelling, stabilization and pruning during development, and similar structural changes can be triggered by learning and changes in sensory experiences^{1–4}. However, the biochemical triggers and mechanisms of *de novo* spine formation in the developing brain and the functional significance of new spines to neuronal connectivity are largely unknown. Here we develop an approach to induce and monitor *de novo* spine formation in real time using combined two-photon laser-scanning microscopy and two-photon laser uncaging of glutamate. Our data demonstrate that, in mouse cortical layer 2/3 pyramidal neurons, glutamate is sufficient to trigger *de novo* spine growth from the dendrite shaft in a location-specific manner. We find that glutamate-induced spinogenesis requires opening of NMDARs (*N*-methyl-D-aspartate-type glutamate receptors) and activation of protein kinase A (PKA) but is independent of calcium-calmodulin-dependent kinase II (CaMKII) and tyrosine kinase receptor B (TrkB) receptors. Furthermore, newly formed spines express glutamate receptors and are rapidly functional such that they transduce presynaptic activity into postsynaptic signals. Together, our data demonstrate that early neural connectivity is shaped by activity in a spatially precise manner and that nascent dendrite spines are rapidly functionally incorporated into cortical circuits.

During postnatal development, the formation and elimination of glutamatergic synapses are thought to be reflected in the growth and retraction of dendritic spines. In cortical pyramidal neurons, waves of new spine growth (spinogenesis) and synapse formation (synaptogenesis) occur at specific developmental stages, followed by pruning as the brain matures⁵. Many signals have been proposed to trigger and regulate *de novo* spine growth in a developing circuit including neurotrophins, neurotransmitters and cell-adhesion molecules^{6–9}. To uncover the triggers for and mechanisms of spinogenesis, we imaged dendrites of enhanced green fluorescent protein (EGFP)-expressing cortical layer 2/3 pyramidal neurons while releasing glutamate at a specific dendritic location by two-photon laser-induced photolysis of (4-methoxy-7-nitroindolyl)-glutamate (MNI-glutamate) (Fig. 1). Analysis was performed in acute cortical brain slices from young mice (postnatal day (P) 8–12), a period in which spinogenesis occurs *in vivo*¹⁰.

Stimulation near the edge of a dendrite with 40 0.5-ms laser pulses at 0.5 Hz in a Mg^{2+} -free extracellular solution induced growth of a new spine in approximately 14% of cases (Fig. 1a–d and Supplementary Fig. 1), showing the possibility of *de novo* spinogenesis induced by glutamate exposure¹¹. Increasing stimulation frequency and laser pulse duration while maintaining the total number of stimuli at 40 increased the rate of spinogenesis such that, at 5 Hz with 4 ms duration, a maximal success rate of approximately 50% was achieved (Fig. 1c). Nascent spines arose from the dendrite where glutamate was released with high specificity (Fig. 1b) such that more than 70% of them grew within 1 μm of the uncaging spot (Fig. 1e) and 94% of them grew on the side of the dendrite exposed to glutamate.

In 128 of 132 examples of glutamate-induced spinogenesis, the spine was seen to emerge without a filopodial stage (see Supplementary Fig. 2a for an exception). Instead, spine growth occurred incrementally but explosively such that the spine head volume increased from 10 to 90% of maximum within 11.8 ± 1.5 pulses of glutamate (5.9 ± 0.8 s at 2 Hz stimulation) (Fig. 2a–c and Supplementary Fig. 3). The final sizes and lengths of the newborn spines were heterogeneous but not different from those of pre-existing neighbouring spines (Fig. 2d, e). The lifetime of newly formed spines was variable such that approximately 20% lasted less than 2 min but those that lasted 5 min were stable and remained for at least 30 min (Supplementary Fig. 4). Thus, these newly formed spines either did not require continued exposure to glutamate for maintenance or they received glutamate from an alternative source such as an axonal bouton.

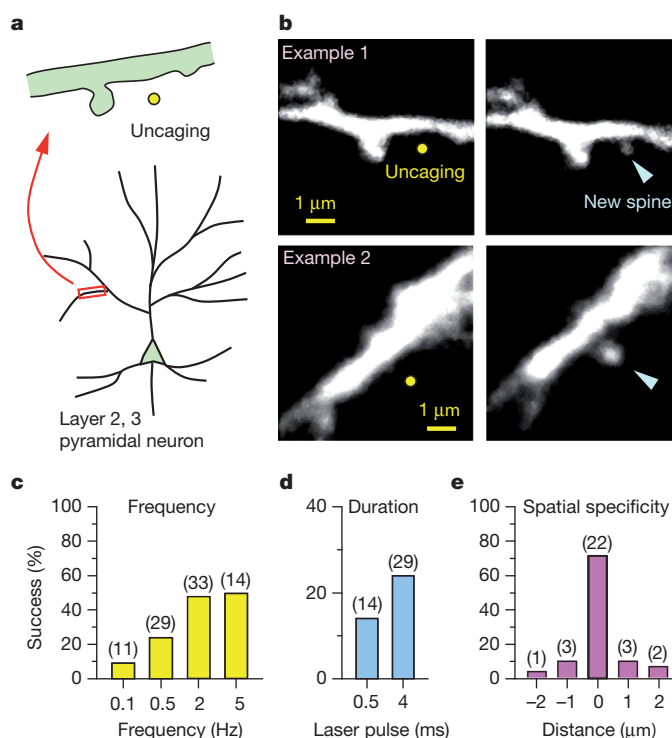


Figure 1 | *De novo* spine generation is induced by glutamate uncaging. **a**, Dendrites of EGFP-expressing neurons in acute slices from P8–12 mice were visualized with two-photon laser scanning microscopy, and glutamate was released by photolysis of caged glutamate near a low-spine density section of dendrite. **b**, Examples of *de novo* spine formation induced by photolytic release of glutamate (40 pulses of MNI-glutamate uncaging at 2 Hz in Mg^{2+} -free artificial cerebrospinal fluid). Yellow circles, the uncaging spots; arrowheads, new spines. **c–e**, Most new spines grew near the uncaging spot and the success percentage depended on the frequency (**c**, laser pulse duration = 4 ms) and duration (**d**, stimulation frequency = 0.5 Hz) of glutamate uncaging. Experiment numbers for each bar are indicated in parentheses.

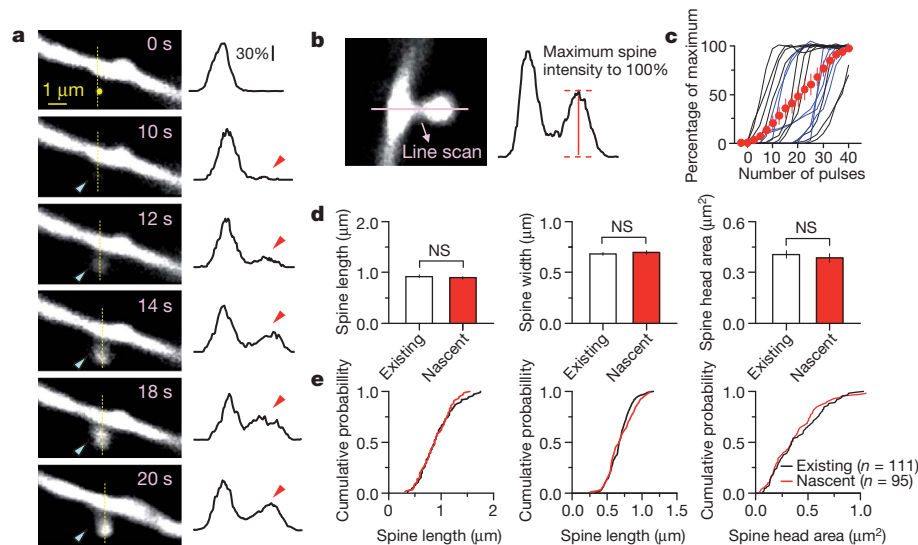


Figure 2 | New spines grow rapidly and acquire morphology similar to pre-existing spines. **a**, Left, time-lapse images of spine formation during glutamate uncaging (40 pulses, 2 Hz) showing the uncaging spot (yellow circle) and nascent spine (blue arrowhead). Right, fluorescence intensity profiles along the yellow line reveal that the spine head fluorescence increases gradually but rapidly (red arrowhead). **b**, Illustration of the measurement of spine head fluorescence during spinogenesis as a percentage of the maximum fluorescence intensity reached. **c**, Time course of individual (black, 2 Hz; blue, 0.5 Hz) and

average (red) fluorescence intensity increases during spinogenesis ($n = 17$). Error bars, s.e.m. **d**, Average of apparent spine length, width and head area from nascent ($n = 95$) and neighbouring existing ($n = 111$) spines (existing and nascent: length: $0.92 \pm 0.03 \mu\text{m}$, $0.89 \pm 0.03 \mu\text{m}$, $P > 0.1$; width: $0.68 \pm 0.02 \mu\text{m}$, $0.70 \pm 0.02 \mu\text{m}$, $P > 0.1$; head area: $0.41 \pm 0.02 \mu\text{m}^2$, $0.38 \pm 0.03 \mu\text{m}^2$, $P > 0.1$). **e**, Cumulative distributions demonstrating that the morphologies of pre-existing and nascent spines are not different.

Glutamate-induced spinogenesis was restricted within postnatal developmental such that its efficiency diminished by P14–15 and it failed to occur by P19–20 (Supplementary Fig. 5). This was not due to decreased glutamate receptor activation in older animals because the uncaging-evoked excitatory postsynaptic current (uEPSC) was larger at P19–20 than at P10–12 (Supplementary Fig. 6).

Previous ultrastructural studies have revealed a high frequency of dendrite shaft synapses in hippocampus in early postnatal life that decreases as spinogenesis occurs¹², leading to a model of synaptic development in which synapses are initially formed directly onto the dendritic shaft and a spine subsequently grows from this point with the synapse attached. On the other hand, rapid movement of a physically connected spine head and axonal bouton together through a complex neuropile is difficult to reconcile with the high density of crossing axons and dendrites¹³. Our data demonstrate that glutamate uncaging-induced spinogenesis occurs with high spatial specificity and probability on the side of the dendrite exposed to glutamate. These findings place a lower limit of approximately $1 \mu\text{m}^{-1}$ for the density of dendritic shaft synapses required to support this model of spinogenesis. To estimate the number of dendritic shaft synapses, Ca^{2+} imaging was performed in conditions in which most synapses formed onto a stretch of dendrite were activated (approximately 90%, Supplementary Fig. 7). Under these conditions, we observed hotspots of Ca^{2+} influx in spineless stretches of dendrite at a density of $0.05 \mu\text{m}^{-1}$. This corresponds to a density of dendritic shaft synapses containing NMDARs that is approximately 20-fold less than necessary to explain the specificity and efficiency of glutamate-induced spinogenesis (see Supplementary Fig. 7 for further discussion).

The high success rate of spinogenesis induced by glutamate uncaging allows identification of the signalling pathways that couple activity to spine growth (Fig. 3). Previous analyses of spine generation induced by electrical stimulation indicate a requirement for NMDARs in this process^{14,15}, and at this age NMDARs are found throughout the dendrite (Supplementary Fig. 8). Preventing NMDAR activation with the antagonist 3-(2-carboxypiperazin-4-yl)propyl-1-phosphonic acid (CPP) nearly abolished spinogenesis whereas it was unaffected by inhibiting α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA)/

kainate glutamate receptors with 2,3-dihydroxy-6-nitro-7-sulfamoylbenzo[f]quinoxaline-2,3-dione (NBQX) (Fig. 3a). The voltage-gated sodium channel antagonist tetrodotoxin also did not affect spinogenesis, discarding the possibility that postsynaptic action potentials are necessary. Addition of extracellular Mg^{2+} significantly decreased the success rate, suggesting that the degree of current flux through NMDARs plays a crucial role in triggering spine formation. Blocking either mGluR1 or mGluR5 using 7-hydroxyiminocyclopropan[b]chromen-1a-carboxylic acid ethyl ester (CPCCOEt) or 2-methyl-6-(phenylethynyl)-pyridine (MPEP) or with the less selective group I mGluR antagonist 1-aminoinidan-1,5-dicarboxylic acid (AIDA) demonstrated that neither was strictly necessary for spinogenesis. Lastly, depleting Ca^{2+} stores with cyclopiazonic acid or thapsigargin significantly inhibited spine formation. Thus our data indicate that NMDARs, with additional contributions from intracellular stores provide, the coupling between glutamate and activation of intracellular pathways responsible for spinogenesis.

We further considered intracellular signalling pathways that might be activated in a spatially delimited fashion by glutamate and could provide the spatial information necessary for local spine growth. Previous studies of long-term potentiation and associated spine enlargement in older animals demonstrated a need for activation of CaMKII or signalling by neurotrophin receptor tyrosine kinases such as the BDNF receptor TrkB^{16–19}. However, we found that activity-dependent spinogenesis is unaffected by the kinase inhibitors KN-62, KN-93 and K252a, indicating independence from CaMKII and TrkB signalling (Fig. 3a).

The cyclic AMP (cAMP)-activated kinase PKA is required for long-term potentiation induction in younger neurons²⁰, and gradients in its concentration can be maintained over micrometre-length scales²¹. Raising cAMP concentration by applying the adenylate cyclase activator forskolin was not sufficient to generate spines on its own ($n = 3$, data not shown), but glutamate uncaging in its presence increased the spinogenesis success rate to approximately 80% (Fig. 3a). In the presence of forskolin, multiple spines often grow with each induction attempt (Fig. 3b, c), including at sites more distant from the uncaging location (Fig. 3d). In addition, the PKA inhibitor H-89 prevented new

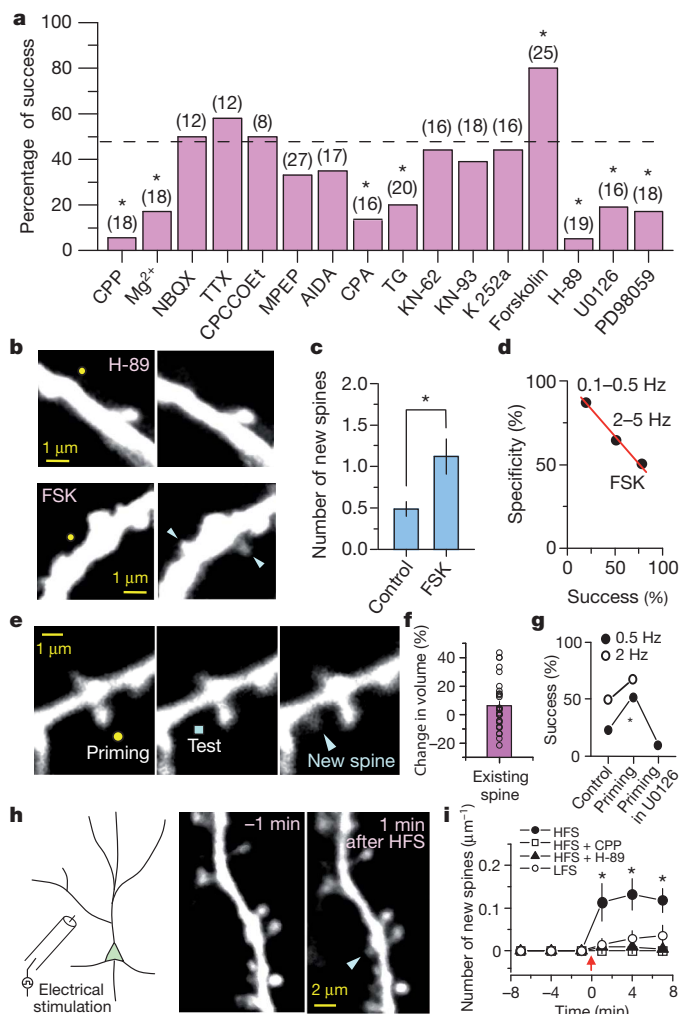


Figure 3 | Molecular mechanisms of glutamate-induced spine formation. **a**, Spine formation using the 40 pulses at 2 Hz protocol was tested in the presence of pharmacological agents. The dotted line indicates the success percentage in control conditions with which statistical comparison was made. The number of induction attempts for each condition is given in parentheses, and the numbers of successes and total trials are summarized in Supplementary Table 2. TG, thapsigargin; CPA, cyclopiazonic acid. **b**, Examples of blockade of spine generation and of exuberant spine growth. Images taken before (left) and after (right) glutamate uncaging in the presence of either H-89 or forskolin (FSK). Arrowheads, nascent spines. **c**, Summary of the average number of new spines with each induction attempt (control: 0.48 ± 0.09 , $n = 33$; FSK: 1.12 ± 0.21 , $n = 25$, $P < 0.005$). **d**, Inverse relation between the location specificity and success rate. Data from three groups (0.1–0.5 Hz, 2–5 Hz and FSK at 2 Hz) are plotted. Specificity was measured as the percentage of cases in which the spine arose within 1 μm of the uncaging spot. **e**, Representative images of priming experiments in which 40 pulses of 2 Hz glutamate uncaging were delivered to a pre-existing spine (yellow circle) followed by an additional 40 pulses (0.5 Hz or 2 Hz) delivered to the nearby dendrite (blue square). Releasing glutamate did not cause enlargement of the pre-existing spine head (left), but did trigger new spine growth from the dendrite (middle, right). **f**, Changes in the fluorescence of pre-existing spine heads exposed to the priming stimulus (individual spines: circles; bar graph: average \pm s.e.m.) ($n = 31$). **g**, Percentage of successful spine generation at the indicated test frequencies with and without priming. U0126 prevented spinogenesis facilitated by the priming protocol. **h**, The experiment (left) and images of a dendrite 1 min before (middle) and after (right) HFS (2×100 pulses at 100 Hz). **i**, Average numbers of new spines generated by HFS per micrometre of dendrite in control conditions (at 7 min: 0.12 ± 0.03 , $n = 11$, $P < 0.005$), in the presence of CPP (0, $n = 8$, $P > 0.5$), or H-89 (0.004 ± 0.001 , $n = 12$; $P > 0.5$). The same number of pulses at a lower frequency (LFS, 10 Hz) generated fewer new spines (0.035 ± 0.024 , $n = 7$, $P > 0.1$). Error bars, s.e.m.

spine growth (Fig. 3a), indicating that PKA activity is necessary but not sufficient for spinogenesis.

The small guanosine triphosphatase Ras is activated by Ca^{2+} influx through NMDARs and signals through mitogen-activated protein kinase (MAPK) to promote long-term potentiation^{17,22}. We found that the MAPK pathway was necessary for spinogenesis because the success rate was significantly reduced by blocking the upstream activator MAPK kinase 1/2 (MEK1/2) with U0126 or MEK1 with PD98059 (Fig. 3a). In pyramidal neurons, activated Ras diffuses from active spines to neighbouring spines and heterosynaptically facilitates plasticity¹⁷. To examine if a similar phenomenon potentiates spinogenesis, we delivered a 'priming' stimulus to a pre-existing spine (40 pulses at 2 Hz, Fig. 3e) and then a 'test' stimulus to the dendritic shaft within 1–2 min. In contrast to previous studies of older neurons in hippocampus^{16–18}, we found no consistent increase in volume of the existing spine in response to this priming stimulus (Fig. 3f), supporting the idea that spinogenesis is not simply due to the growth of an undetectable pre-existing spine. Nevertheless, the success rate of spine generation was enhanced by the priming stimulus in a MEK1/2-dependent manner such that the low level of spinogenesis evoked by a 0.5 Hz test stimulus was increased. In contrast, the priming stimulus did not increase the success rate when the test stimulus was delivered at 2 Hz (Fig. 3g), indicating that the priming protocol shifted the induction threshold for spinogenesis.

With the exception of the NMDAR and PKA antagonists, none of the pharmacological manipulations that altered spinogenesis rates affected dendritic currents or Ca^{2+} transients (Supplementary Fig. 8). As expected, CPP largely abolished the Ca^{2+} transient and the prolonged phase of the currents. Consistent with a facilitation of Ca^{2+} influx through NMDARs by PKA^{23,24}, H-89 lowered dendritic Ca^{2+} transients by approximately 20% (Supplementary Table 1). However, this effect is insufficient to explain the near abolition of spinogenesis because the rate of spinogenesis evoked by 0.5 ms uncaging pulses in control conditions was higher than that evoked by 4 ms pulses in H-89, despite eliciting smaller dendritic Ca^{2+} transients (Supplementary Fig. 8 and Supplementary Table 1).

We examined whether endogenous synaptic activity generates new spines in cortical tissue from young mice through similar pathways. In normal extracellular Mg^{2+} , high-frequency (100 pulses at 100 Hz, delivered twice, separated by 10 s) but not low-frequency (10 Hz) electrical tetani rapidly triggered new spine growth (Fig. 3h, i). Blockade of NMDARs or PKA prevented this synaptically evoked spine growth (Fig. 3i) as well as spontaneous new spine growth that occurred when NMDAR activation was increased by removing extracellular Mg^{2+} and blocking inhibitory GABA_A receptors (Supplementary Fig. 9). Hence, multiple experimental models demonstrate that activity-dependent spinogenesis in developing cortex requires NMDAR- and PKA-dependent signalling.

To determine if nascent spines detect synaptically released glutamate and are functionally incorporated into the circuit, we generated new spines by high-frequency stimulation (HFS) and examined their synaptic responses using optical quantal analysis of synaptic properties. A whole-cell recording was obtained and the probability and amplitude of synaptically evoked NMDAR-mediated Ca^{2+} transients in the spine head were monitored (Fig. 4a). Using the Ca^{2+} -sensitive green fluorophore Fluo-5F, we detected stimulus-evoked Ca^{2+} transients in the heads of newly grown spines (Fig. 4b, c), demonstrating that they sense synaptic activity within a neural circuit within 30 min after growth. Similar results were obtained in five of seven new spines and in 11 of 16 pre-existing spines. Analysis of the spines in which an evoked Ca^{2+} transient could be detected indicated that nascent spines displayed smaller and less frequent synaptically evoked Ca^{2+} transients than pre-existing spines (Fig. 4d, e). Similar results were obtained for new spines that grew in response to glutamate uncaging and were probed using a glass-stimulating electrode placed near the spine (Supplementary Fig. 10).

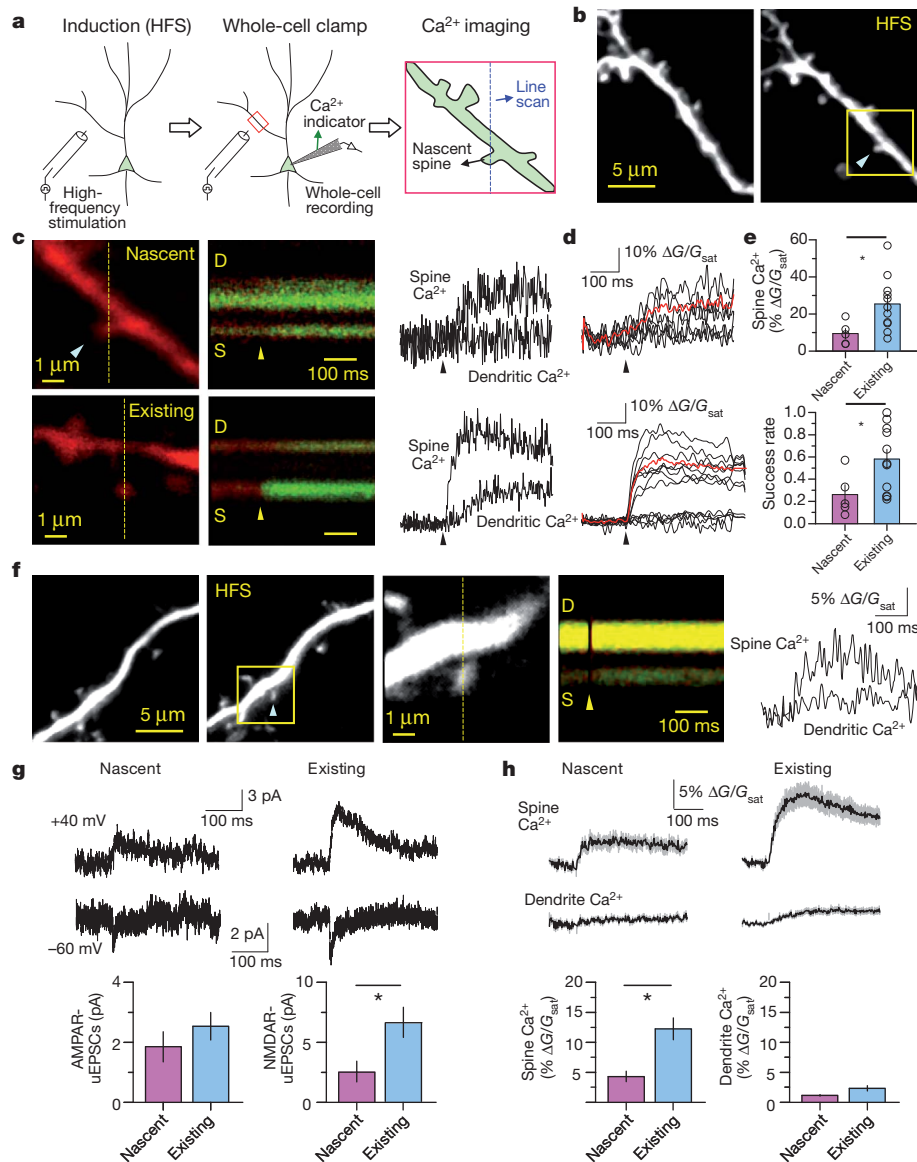


Figure 4 | Functional characterization of new spines. **a**, Schematic of the experimental procedure. HFS was delivered approximately 30 μm from the target dendritic region (red box). After nascent spines were identified, Ca^{2+} indicator was loaded into the cell through a whole-cell recording pipette. The newly generated spine was examined at higher temporal and spatial resolution to measure synaptically evoked Ca^{2+} transients in the spine head and perform optical quantal analysis. **b**, Images before and after HFS showing the new spine (arrowhead) and the area subsequently analysed at higher resolution (yellow box). **c**, Images (left) of newly generated (top) and pre-existing (bottom) spines filled with the fluorophore Alexa 594 (red, 20 μM) and Fluo-5F (green, 300 μM). Fluorescence was collected (middle) and quantified (right) from a line-scan intersecting the spine (S) and dendrite (D) after electrical stimulation (arrowhead). The increases in green fluorescence indicate Ca^{2+} entry. **d**, Green fluorescence transients collected in consecutive trials (black) showing successes and failures. The average 'success' fluorescence transient is also shown (red).

The AMPA receptor (AMPA) and NMDAR content of new spines was characterized using glutamate uncaging after HFS-induced spinogenesis. In all cases, we detected fast uEPSCs at a holding potential of -60 mV and large prolonged uEPSCs at $+40$ mV (Fig. 4g), consistent with AMPAR- and NMDAR-mediated currents previously characterized in these cells²⁵. Similarly, uncaging-evoked Ca^{2+} transients were clearly visible at -60 mV (Fig. 4h) but were larger at $+40$ mV (data not shown), consistent with the known properties of NMDAR-mediated Ca^{2+} influx. Despite similar-sized AMPAR-uEPSCs in new and pre-existing spines, NMDAR-uEPSCs and Ca^{2+} influx were significantly lower in nascent

than pre-existing spines (Fig. 4g, h), similar to previous descriptions of spontaneously appearing new spines in hippocampal organotypic slices²⁶. Therefore the smaller Ca^{2+} transients measured in nascent spines by synaptic activation probably reflect a smaller number of postsynaptic NMDARs.

In this study, we established a protocol for the reliable and spatio-temporally precise induction of spinogenesis. These experiments demonstrate that glutamate is sufficient to trigger rapid spine formation and suggest that neurons use glutamate release to establish circuit wiring. Thus these data support the hypothesis that axonal growth and

glutamate release may be the triggering event in synapse formation such that axonal bouton localization is an important early step for precise neuronal circuit formation^{10,27}. Given the involvement of NMDARs, Ca^{2+} stores, cAMP, PKA and MAPK in activity-dependent spinogenesis, it is likely that many neuromodulators that regulate these molecules may influence the capacity or threshold for new spine formation. For instance, activation of dopaminergic, serotonergic or adrenergic receptors that signal by $\text{G}\alpha_s$ may facilitate spinogenesis, whereas receptors that activate $\text{G}\alpha_i$ -coupled signalling may function as inhibitory signals.

Lastly, we provide experimental evidence that spines that grow *de novo* in developing cortical tissue become rapidly functionally integrated into the circuit such that they sense synaptically released glutamate through AMPARs and NMDARs. Whether these nascent spines are rapidly physically associated with a presynaptic bouton and display the ultrastructural correlates of a synapse is unknown^{26,28–30}. Our results indicate that spines can grow *de novo* without the need for a filopodial intermediate and probably without a dendritic-shaft synapse stage. In total, this study demonstrates that synaptic activity can rapidly modify neuronal connectivity with high accuracy by generating new circuit elements.

METHODS SUMMARY

All procedures on animals followed protocols approved by the Harvard Standing Committee on Animal Care and National Institutes of Health guidelines. *In utero* electroporation of EGFP was performed at embryonic day 15.5 in C57BL/6 mice. All studies were performed on layer 2/3 pyramidal neurons in acute coronal slices identified by their characteristic morphology, position in the slice and expression of GFP. Two-photon glutamate uncaging and imaging was performed using custom microscopes. To induce spine growth, 40 uncaging pulses were delivered at varying frequencies to a spot approximately 0.5 μm from the edge of the dendrite. Synaptically induced spine growth was triggered with two high-frequency stimuli (100 pulses at 100 Hz) separated by 10 s delivered by a bipolar electrode positioned approximately 30 μm from the target dendrite. For Ca^{2+} imaging, neurons were loaded through the whole-cell recording electrode with Alexa Fluor-594 (20 μM) and Fluo-5F (300 μM) and the amplitudes of fluorescence transients were quantified as a fraction of the maximal green fluorescence achieved in saturating (sat) Ca^{2+} concentrations (G_{sat}). For optical quantal analysis, the synapse associated with a visualized spine was stimulated using a closely positioned glass electrode. The position of the electrode and stimulus intensity were adjusted until (1) Ca^{2+} transients were evoked in the spine head that demonstrated stochastic failures and successes, and (2) Ca^{2+} transients in other spines and the dendritic shaft in the field of view were not evoked. Fisher's exact test was used to compare the efficacy of spinogenesis across conditions. For each spinogenesis trial, an observer blind to the experimental condition was asked to identify if (1) a new spine had grown and (2), if so, how many spines had grown.

Received 25 October 2010; accepted 4 March 2011.

Published online 8 May 2011.

- Grutzendler, J., Kasthuri, N. & Gan, W. B. Long-term dendritic spine stability in the adult cortex. *Nature* **420**, 812–816 (2002).
- Hofer, S. B., Mrsic-Flogel, T. D., Bonhoeffer, T. & Hubener, M. Experience leaves a lasting structural trace in cortical circuits. *Nature* **457**, 313–317 (2009).
- Trachtenberg, J. T. *et al.* Long-term *in vivo* imaging of experience-dependent synaptic plasticity in adult cortex. *Nature* **420**, 788–794 (2002).
- Zuo, Y., Yang, G., Kwon, E. & Gan, W. B. Long-term sensory deprivation prevents dendritic spine loss in primary somatosensory cortex. *Nature* **436**, 261–265 (2005).
- Rakic, P., Bourgeois, J. P., Eckenhoff, M. F., Zecevic, N. & Goldman-Rakic, P. S. Concurrent overproduction of synapses in diverse regions of the primate cerebral cortex. *Science* **232**, 232–235 (1986).
- Dalva, M. B., McClelland, A. C. & Kayser, M. S. Cell adhesion molecules: signalling functions at the synapse. *Nature Rev. Neurosci.* **8**, 206–220 (2007).

- Scheiffele, P. Cell-cell signaling during synapse formation in the CNS. *Annu. Rev. Neurosci.* **26**, 485–508 (2003).
- Sudhof, T. C. Neuroligins and neuroligins link synaptic function to cognitive disease. *Nature* **455**, 903–911 (2008).
- Yuste, R. & Bonhoeffer, T. Genesis of dendritic spines: insights from ultrastructural and imaging studies. *Nature Rev. Neurosci.* **5**, 24–34 (2004).
- Miller, M. & Peters, A. Maturation of rat visual cortex. II. A combined Golgi-electron microscope study of pyramidal neurons. *J. Comp. Neurol.* **203**, 555–573 (1981).
- Richards, D. A. *et al.* Glutamate induces the rapid formation of spine head protrusions in hippocampal slice cultures. *Proc. Natl Acad. Sci. USA* **102**, 6166–6171 (2005).
- Fiala, J. C., Feinberg, M., Popov, V. & Harris, K. M. Synaptogenesis via dendritic filopodia in developing hippocampal area CA1. *J. Neurosci.* **18**, 8900–8911 (1998).
- Sorra, K. E., Fiala, J. C. & Harris, K. M. Critical assessment of the involvement of perforations, spinules, and spine branching in hippocampal synapse formation. *J. Comp. Neurol.* **398**, 225–240 (1998).
- Engert, F. & Bonhoeffer, T. Dendritic spine changes associated with hippocampal long-term synaptic plasticity. *Nature* **399**, 66–70 (1999).
- Maletic-Savatic, M., Malinow, R. & Svoboda, K. Rapid dendritic morphogenesis in CA1 hippocampal dendrites induced by synaptic activity. *Science* **283**, 1923–1927 (1999).
- Harvey, C. D. & Svoboda, K. Locally dynamic synaptic learning rules in pyramidal neuron dendrites. *Nature* **450**, 1195–1200 (2007).
- Harvey, C. D., Yasuda, R., Zhong, H. & Svoboda, K. The spread of Ras activity triggered by activation of a single dendritic spine. *Science* **321**, 136–140 (2008).
- Matsuzaki, M., Honkura, N., Ellis-Davies, G. C. & Kasai, H. Structural basis of long-term potentiation in single dendritic spines. *Nature* **429**, 761–766 (2004).
- Tanaka, J. *et al.* Protein synthesis and neurotrophin-dependent structural plasticity of single dendritic spines. *Science* **319**, 1683–1687 (2008).
- Yasuda, H., Barth, A. L., Stellwagen, D. & Malenka, R. C. A developmental switch in the signaling cascades for LTP induction. *Nature Neurosci.* **6**, 15–16 (2003).
- Zaccolo, M. & Pozzan, T. Discrete microdomains with high concentration of cAMP in stimulated rat neonatal cardiac myocytes. *Science* **295**, 1711–1715 (2002).
- Di Cristo, G. *et al.* Requirement of ERK activation for visual cortical plasticity. *Science* **292**, 2337–2340 (2001).
- Chalifoux, J. R. & Carter, A. G. GABA_B receptors modulate NMDA receptor calcium signals in dendritic spines. *Neuron* **66**, 101–113 (2010).
- Skeberdis, V. A. *et al.* Protein kinase A regulates calcium permeability of NMDA receptors. *Nature Neurosci.* **9**, 501–510 (2006).
- Busetto, G., Higley, M. J. & Sabatini, B. L. Developmental presence and disappearance of postsynaptically silent synapses on dendritic spines of rat layer 2/3 pyramidal neurons. *J. Physiol. (Lond.)* **586**, 1519–1527 (2008).
- Zito, K., Scheuss, V., Knott, G., Hill, T. & Svoboda, K. Rapid functional maturation of nascent dendritic spines. *Neuron* **61**, 247–258 (2009).
- Hamori, J. The inductive role of presynaptic axons in the development of postsynaptic spines. *Brain Res.* **62**, 337–344 (1973).
- De Roo, M., Klausner, P., Mendez, P., Poggio, L. & Muller, D. Activity-dependent PSD formation and stabilization of newly formed spines in hippocampal slice cultures. *Cereb. Cortex* **18**, 151–161 (2008).
- Nagerl, U. V., Kostinger, G., Anderson, J. C., Martin, K. A. & Bonhoeffer, T. Protracted synaptogenesis after activity-dependent spinogenesis in hippocampal neurons. *J. Neurosci.* **27**, 8149–8156 (2007).
- Knott, G. W., Holtmaat, A., Wilbrecht, L., Welker, E. & Svoboda, K. Spine growth precedes synapse formation in the adult neocortex *in vivo*. *Nature Neurosci.* **9**, 1117–1124 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Sabatini laboratory for their comments on the manuscript and assistance with data analysis. We are grateful to S. Nazia for technical support and for acting as the blind evaluator. This work was supported by a SFARI grant from the Simons Foundation and the National Institute of Neurological Disorders and Stroke (NS046579).

Author Contributions H.B.K. and B.L.S. designed the experiments and wrote the paper. H.B.K. performed all the experiments, analysed the data (other than spine counting by a blind, third-party observer) and prepared the figures.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.L.S. (bsabatini@hms.harvard.edu).

UBCH7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids

Dawn M. Wenzel¹, Alexei Lissounov¹, Peter S. Brzovic¹ & Rachel E. Klevit¹

Although the functional interaction between ubiquitin-conjugating enzymes (E2s) and ubiquitin ligases (E3s) is essential in ubiquitin (Ub) signalling, the criteria that define an active E2–E3 pair are not well established. The human E2 UBCH7 (also known as UBE2L3) shows broad specificity for HECT-type E3s¹, but often fails to function with RING E3s *in vitro* despite forming specific complexes^{2–4}. Structural comparisons of inactive UBCH7–RING complexes with active UBCH5–RING complexes reveal no defining differences^{3,4}, highlighting a gap in our understanding of Ub transfer. Here we show that, unlike many E2s that transfer Ub with RINGs, UBCH7 lacks intrinsic, E3-independent reactivity with lysine, explaining its preference for HECTs. Despite lacking lysine reactivity, UBCH7 exhibits activity with the RING-in-between-RING (RBR) family of E3s that includes parkin (also known as PARK2) and human homologue of ariadne (HHARI; also known as ARIH1)^{5,6}. Found in all eukaryotes⁷, RBRs regulate processes such as translation⁸ and immune signalling⁹. RBRs contain a canonical C3HC4-type RING, followed by two conserved Cys/His-rich Zn²⁺-binding domains, in-between-RING (IBR) and RING2 domains, which together define this E3 family⁷. We show that RBRs function like RING/HECT hybrids: they bind E2s via a RING domain, but transfer Ub through an obligate thioester-linked Ub (denoted ~Ub), requiring a conserved cysteine residue in RING2. Our results define the functional cadre of E3s for UBCH7, an E2 involved in cell proliferation¹⁰ and immune function¹¹, and indicate a novel mechanism for an entire class of E3s.

RING and U-box E3s facilitate Ub transfer directly from an activated E2~Ub to a lysine on a target protein. Therefore, E2s that function with RINGs must be catalytically competent to form an isopeptide bond between Ub and lysine. Previous characterization of E2 activity demonstrates that some E2s can transfer Ub to free lysine independent of an E3 (ref. 12), providing a framework to examine E2 function. We compared the intrinsic, E3-independent reactivity of UBCH7~Ub and UBCH5C~Ub with free amino acids that represent Ub acceptors: lysine, serine, threonine¹³, cysteine¹⁴, or arginine as a control (Fig. 1a). UBCH5C~Ub reacts completely with either cysteine or lysine, but not other amino acids, indicating the side-chain

functional group is the relevant nucleophile. Notably, UBCH7 reacts only with cysteine. Reaction time courses for UBCH7~Ub and UBCH5C~Ub with free lysine show that UBCH5C~Ub is nearly depleted after 15 min, while after 60 min UBCH7~Ub shows no detectable reaction (Fig. 1b). The lack of reactivity of UBCH7 is lysine-specific and cannot be attributed to UBCH7~Ub being intrinsically more stable, as both E2s react equally rapidly with cysteine (Supplementary Fig. 2a). UBE2K and UBC13, E2s known to function with RINGs^{15,16}, both react with cysteine and lysine (Supplementary Fig. 2b), indicating that lysine reactivity is a general feature of RING-active E2s. The reactivity properties of UBCH7 are conserved, as the *Caenorhabditis elegans* orthologue Ubc18 (ref. 17) also lacks lysine reactivity (Supplementary Fig. 3).

To determine which residues in E2s are important for lysine reactivity, the active site sequences of lysine-reactive E2s were aligned with that of UBCH7 (Fig. 2a). Two residues in UBCH7 are distinctly different: D87 and D117 (in UBCH5C numbering) are proline and histidine, respectively, in UBCH7. To establish whether these residues contribute to lysine reactivity, each was mutated in UBCH5C and lysine reactivity was measured (Fig. 2b). The effect of substitution at position 87 ranges from no effect for the isosteric mutation D87N to complete loss of lysine reactivity for the charge-swapped D87K mutation. UBCH5C(D87E) and UBCH5C(D87P) have intermediate reactivities. Consistent with D87 having a general role in lysine reactivity, mutation of the analogous residue in UBE2K(D94E) results in decreased lysine reactivity and impaired formation of free poly-Ub chains (Supplementary Fig. 4a, b). Substitution of D117 in UBCH5C with a histidine as found in UBCH7 greatly decreases lysine reactivity (Fig. 2b). A structurally analogous residue in the SUMO E2 (D127 in UBC9), has been shown to lower the pK_a of a lysine approaching the active site¹⁸. The invariant active-site asparagine residue (N77 in UBCH5C) is recognized for its role in isopeptide catalysis¹⁹, and its mutation to serine abolishes UBCH5C~Ub lysine reactivity in our assay. Identification of several residues that affect the intrinsic lysine reactivity of E2s indicates that the determinants are probably multifactorial. Accordingly, we failed to convert UBCH7 into a lysine-reactive E2 by mutation (Supplementary Table).

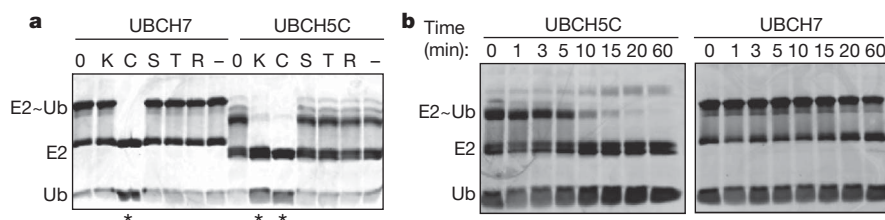


Figure 1 | UBCH7 does not react with free lysine. **a**, Coomassie-stained SDS-PAGE of UBCH7~Ub (left) and UBCH5C~Ub (right) incubated with amino acids lysine, serine, threonine, arginine, or buffer (–). Reactions were quenched in non-reducing loading buffer. Starting amounts of E2~Ub before amino acid addition are indicated as '0'. Reactivity with amino acids is indicated by loss of

E2~Ub and concurrent increase in free Ub (denoted by asterisks) and free E2. **b**, Time-course assays of UBCH5C~Ub and UBCH7~Ub incubated with lysine. Reactions were quenched in non-reducing loading buffer at the indicated times.

¹Department of Biochemistry, Box 357350, University of Washington, Seattle, Washington 98195, USA.

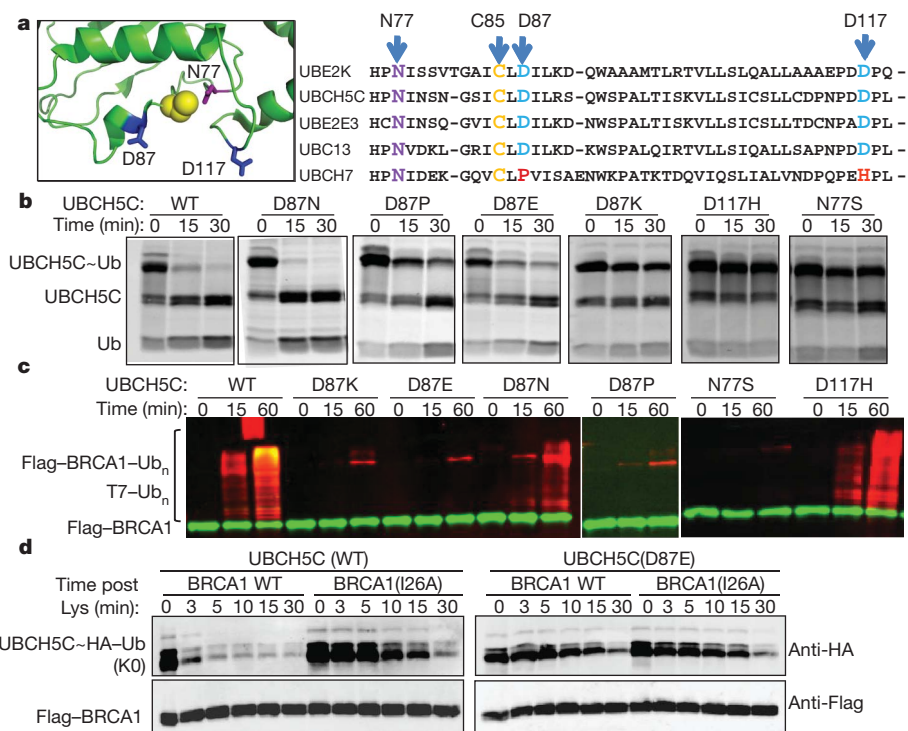


Figure 2 | Lysine reactivity is multifactorial. **a**, Alignment of E2 active-site residues (right). Structure of UBCH5C (green; PDB accession code 2FUH) with active-site residues represented as sticks and active-site cysteine as spheres (left). **b**, Wild-type (WT) or mutant UBCH5C~Ub incubated with lysine as in Fig. 1b. **c**, Western blot of Flag-BRCA1(1–304)/BARD1(26–327) (green) auto-ubiquitination with indicated E2 and T7~Ub (red). Time is measured as

minutes after ATP addition. Multiply ubiquitinated species are indicated as 'Ub_n'. **d**, Same as in Fig. 1b, except a 1:1 (E3:E2) equivalent of Flag-BRCA1(1–112)/BARD1(26–140) was added to either wild-type UBCH5C (left) or the D87E mutant (right) charged with HA~Ub where all lysines have been mutated to arginines (indicated as 'K0'). Reactions were visualized by western blot for HA (Ub(K0)) and Flag (BRCA1) epitopes.

To assess whether the intrinsic reactivity of an E2 is predictive of its functional E3 interaction, UBCH5C mutants were assayed with RING and HECT-type E3s. Of the D87 mutants, only UBCH5C (D87N) (which has wild-type lysine reactivity) is able to function with the RING E3 BRCA1/BARD1 in an auto-ubiquitination assay; the other substitutions are inactive (Fig. 2c). The observed loss of activity results from E2 catalytic defects, as NMR binding experiments confirm that UBCH5C(D87) mutants bind BRCA1 comparably to wild-type UBCH5C (Supplementary Fig. 5). UBCH5C(D117H), which has impaired lysine reactivity, retains some ability to transfer Ub to BRCA1 (Fig. 2c). Given its position at the active site, D117 may provide substrate-specific lysine reactivity as a gating residue. As expected, UBCH5C(N77S) is inactive with BRCA1. Previous studies show that mutation of E2 residue N77 abolishes Ub transfer with RING-type ligases MDM2 (ref. 19), RMA1 (also known as RNF5) (ref. 20), CNOT4 (CCR4-NOT transcription complex subunit 4) and APC/C (anaphase promoting complex/cyclosome) (ref. 21) but retains activity with HECT-type ligases E6-AP (also known as UBE3A), KIAA10 (also known as UBE3C; ref. 19) and NEDD4L (ref. 22). This indicates that the catalytic requirements for trans-thiolation differ from those for isopeptide bond synthesis. Accordingly, the ability to form E3~Ub thioesters with the HECT E3 E6-AP is unaffected in all UBCH5C mutants (Supplementary Fig. 6). Taken together, our results indicate that E2 lysine reactivity is a prerequisite for transfer with RING E3s. Furthermore, lysine-unreactive E2s, such as UBCH5C(N77S), can be diagnostic for differentiating between RING- versus HECT-type Ub transfer mechanisms.

Although many E2s possess an intrinsic reactivity for lysine, most do not transfer Ub to protein substrates independent of an E3. To explore the contribution of E3s to E2 reactivity, we examined the intrinsic reactivity of E2~Ub in the presence of RING or HECT E3s. Interaction with BRCA1/BARD1 (RING) or E6-AP (HECT) does not change the intrinsic reactivity of UBCH5C and UBCH7 with free

amino acids (Supplementary Fig. 7a, b). However, BRCA1/BARD1 enhances UBCH5C~Ub lysine reactivity compared to a reaction containing the E2-binding mutant BRCA1(I26A)^{3,15} (Fig. 2d). Intriguingly, the lysine reactivity of UBCH5C(D87E), which retains both its ability to bind BRCA1/BARD1 and some intrinsic lysine reactivity, is not enhanced by BRCA1 (Fig. 2d). These results indicate that RING binding to a lysine-reactive E2~Ub results in a thioester with enhanced reactivity to lysine and residues such as D87 in UBCH5C couple E3 binding and E2 activation. Notably, E6-AP does not enhance UBCH5C~Ub lysine reactivity (Supplementary Fig. 8b), highlighting a mechanistic difference between RINGs and HECTs. In matched experiments with UBCH7, neither BRCA1/BARD1 nor E6-AP enhance reactivity towards free cysteine (Supplementary Fig. 8a, c).

Besides HECT-type ligases, UBCH7 is reported to function with members of the RING E3 family known as RBRs, which includes parkin and HHARI^{5,6}. This activity runs contrary to our conclusion that UBCH7 lacks lysine reactivity and is consequently restricted to Ub transfer involving trans-thiolation chemistry. Therefore, we examined the Ub ligase mechanism for HHARI and parkin. The minimal ligase (RBR) domains of HHARI_{R1-IBR-R2} (where R1 and R2 are RING1 and 2, respectively) and parkin_{R1-IBR-R2} (see Supplementary Fig. 9 for schematic of constructs) show comparable auto-ubiquitination activity with either UBCH7 or UBCH5C (Fig. 3a, b). In contrast to our results with BRCA1/BARD1, UBCH5C(N77S) exhibits Ub transfer with HHARI_{R1-IBR-R2} and parkin_{R1-IBR-R2} (Fig. 3a, b). As UBCH5C(N77S) activity is restricted to HECT-type ligases, its activity with parkin and HHARI indicates that these ligases do not function via a typical RING-type mechanism.

A hallmark of HECT-type Ub transfer is the formation of an obligate E3~Ub thioester intermediate. Reactions designed to trap a HHARI~Ub conjugate were conducted on ice and quenched with SDS-loading buffer with or without reducing agent (βME) to distinguish reducible thioester-linked Ub from non-reducible isopeptide-linked Ub. A βME-sensitive band corresponding to the molecular

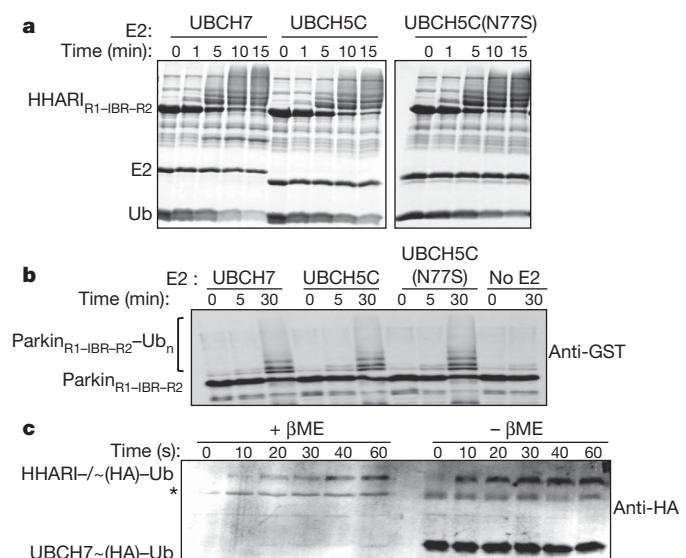


Figure 3 | RBRE E3s function via a HECT-like mechanism. **a**, HHARI_{R1-IBR-R2} auto-ubiquitination assays with the indicated E2 were visualized on a Coomassie-stained reducing SDS gel. Time is measured as minutes after ATP addition. **b**, Auto-ubiquitination assay of parkin_{R1-IBR-R2} with the E2s indicated. Products were visualized by western blotting for the GST tag on parkin. **c**, UBCH7 was pre-charged with HA-Ub and mixed with HHARI_{R1-IBR-R2}. Reactions were quenched in SDS buffer under reducing (+βME) and non-reducing conditions (−βME) and visualized by western blotting for the HA epitope on Ub. A βME-sensitive HA-Ub band corresponding to the molecular weight of HHARI_{R1-IBR-R2}-Ub appears at 10, 20 and 30 s after addition of UBCH7~Ub. Asterisk denotes a cross-reactive band.

weight of HHARI~Ub was detected at 10, 20, and 30 s after addition of pre-charged UBCH7~Ub (Fig. 3c). We next sought to determine the position of the HHARI active-site cysteine. The N-terminal canonical RING1 of HHARI has been shown to be the principal E2-binding region²³, although RING2 is also required for ligase activity⁵. Cysteine 357 in RING2 is highly conserved across RBR ligases (Fig. 4a). C357 is not a Zn²⁺-liganding residue, and mutation of C357 does not destabilize the RING2 structure⁵ (Supplementary Fig. 10), but does abolish the ability of HHARI to transfer Ub⁵. This indicates that C357 may have a catalytic function. Although HHARI_{R1-IBR-R2}(C357A) showed no ligase activity in an auto-ubiquitination assay, HHARI_{R1-IBR-R2}(C357S) generated a single

monoubiquitinated species (Fig. 4b). An analogous oxyester-linked Ub product can be generated on E2s for which the active-site cysteines have been mutated to serines^{20,22,24}. Consistent with its identity as an oxyester (versus isopeptide) bond, the Ub adduct on HHARI_{R1-IBR-R2}(C357S) is labile to alkaline treatment (Supplementary Fig. 11). Formation of an oxyester is unique to C357 as serine substitution of other conserved cysteines (both Zn²⁺- and non-Zn²⁺-liganding) in RING2 does not stall the E3 at a single-Ub adduct but rather impairs (C375S) or abolishes (C367S, C372S) ligase activity (Supplementary Fig. 12). Similarly, the parkin mutant C431S (analogous to HHARI(C357)) eliminates E3 ligase activity (Fig. 4c). We were unable to trap a parkin_{R1-IBR-R2}(C431S) Ub adduct under our reaction conditions. However, we note that the parkin mutation C431F has been consistently shown to abolish parkin's ubiquitination of substrates^{25–27} and genetically predisposes for Parkinson's disease²⁸. The results presented above are not affected by the presence of a GST domain, as several results were reproduced with non-GST versions of HHARI (Supplementary Fig. 13).

HHARI_{R1-IBR-R2}(C357A) effectively binds E2 as GST pull-downs with purified UBCH7 and constructs of GST-HHARI demonstrate that HHARI_{R1-IBR-R2}(C357A) interacts with UBCH7 as efficiently as wild-type HHARI_{R1-IBR-R2} (Fig. 4d). In contrast, a HHARI-RING1 mutant (I188A²³, analogous to BRCA1(I26A)) does not interact detectably with UBCH7 in this assay. Furthermore, C357 is surface accessible and reactive, as wild-type HHARI_{R2} but not HHARI_{R2}(C357A) is readily derivitized by cysteine-modifying reagents (Supplementary Fig. 14).

In the absence of a bona fide substrate, an *in vitro* product of HHARI-catalysed Ub transfer is the non-reducible ubiquitination of UBCH7 (Fig. 4b). Ubiquitination of UBCH7 is E3-dependent, and mutation of HHARI_{R1-IBR-R2}(C357) to serine or alanine abolishes the formation of this product (Fig. 4b). Our finding that HHARI_{R1-IBR-R2}(C357S) forms an oxyester-linked Ub without subsequent transfer of Ub to UBCH7 indicates an ordered mechanism that involves formation of an E3~Ub before modification of UBCH7 (or substrates). Thus, HHARI, unlike other RING E3s, does not facilitate direct transfer from an E2~Ub to a target. We note that the HHARI oxyester-linked Ub conjugate accumulates in low yield compared to the available number of active sites. This is consistent with our failure to observe intrinsic serine reactivity for UBCH7, even at serine concentrations as high as 0.5 M (Fig. 1a and data not shown) and suggests that the unique chemical environment surrounding a target residue (in this case an enzyme active site) contributes to catalysis—a contribution that is absent in the nucleophile

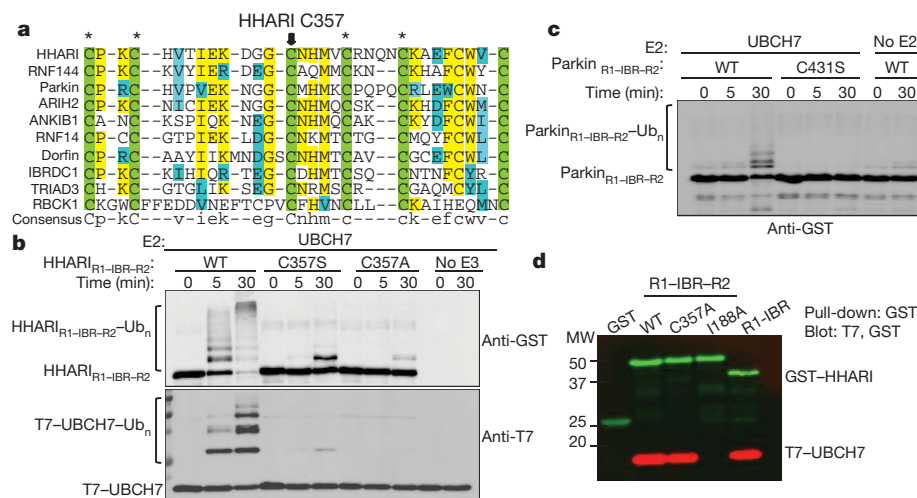


Figure 4 | Cysteine C357 is the active site of HHARI. **a**, Sequence alignment of RING2 domains from human RBR ligases. Conserved cysteine residues are coloured green. Asterisks indicate Zn²⁺-liganding cysteines in the HHARI_{R2} structure⁵. **b**, Auto-ubiquitination assays of T7-tagged UBCH7 and the indicated construct of HHARI. Ubiquitination is measured as time after ATP addition. Reaction products were visualized by western blotting simultaneously for the GST tag on HHARI (top panel) and the T7 tag on UBCH7 (bottom panel). **c**, E3 auto-ubiquitination assay with UBCH7 and the indicated construct of parkin. **d**, GST pull-downs of purified constructs of GST-tagged HHARI and T7-tagged UBCH7. Bound protein was eluted and visualized by western blotting for T7 (red) and GST (green) epitopes simultaneously.

assay. Our combined results are consistent with HHARI and parkin functioning via a HECT-like mechanism whereby RING1 harbours the E2-binding site and RING2 harbours the active-site cysteine.

Our characterization of the reactivity of UBCH7 resulted in two unexpected discoveries. First, whereas E2s known to work with RING-type E3s have E3-independent reactivity towards lysine, the intrinsic reactivity of UBCH7 is restricted to cysteine and, consequently, its Ub transfer activity is restricted to HECT-type ligases. Second, in the process of confirming reports that UBCH7 is active with E3s in the RBR family, we discovered that these RING-containing E3s function like HECTs in that they require an obligate trans-thiolation step during Ub transfer. Both findings have important implications for guiding our understanding of ubiquitination pathways. RING1 of HHARI does not harbour catalytic activity, and neither HHARI_{R1-IBR} nor HHARI_{R2} enhance lysine reactivity of UBCH5C (Supplementary Fig. 15). Our results underscore the diversity of structures that facilitate thiol-based Ub transfer, enzymes that include bacterial HECT-like E3s that bear no homology to eukaryotic HECT counterparts²⁹. Knockdown and over-expression studies indicate that UBCH7 regulates S-phase progression into G2, but neither the E3(s) nor targets responsible have been identified¹⁰. Our results indicate that the relevant E3(s) will be found among HECT or RBRs. Although it is possible that UBCH7 cooperates with RINGs such as BRCA1 to modify substrate cysteines, we have not observed such species.

Among human E2s, only five residue types are found at the position analogous to D87 of UBCH5C: aspartate, serine, asparagine, glutamate and proline. The tolerance for asparagine and serine at position 87 indicates that the negative charge of D87 may not be critical for its role in catalysis, but instead a hydrogen-bonding function seems likely, possibly interacting with the conjugated Ub. Of human E2s, only UBCH7 and UBCH8 have a proline at position 87. Although we did not test UBCH8 for lysine reactivity, it functions primarily with the HECT-type ligase HERC5 to transfer the Ub-like protein ISG15 to substrates³⁰. Like UBCH7, we anticipate UBCH8 activity will be limited to HECT or RBR E3s. Our effort to understand the mechanism of UBCH7-mediated Ub transfer highlights the predictive power of elucidating E2 mechanisms to understand the E3s with which they function.

METHODS SUMMARY

Plasmids, protein expression/purification, E3 auto-ubiquitination assays and NMR experiments were performed as described previously^{3,15}. Modifications and details are described in Methods. For intrinsic reactivity assays, E2s were charged with Ub for 20–30 min at 37 °C before addition of cysteine, arginine, lysine, serine, threonine or buffer (final concentration 50 mM, pH 7.0). After 15–20 min, reactions were quenched in non-reducing loading-buffer and visualized by Coomassie-stained SDS-PAGE. Reactivity time courses with lysine and cysteine were performed similarly with samples quenched at the indicated times.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 October 2010; accepted 23 February 2011.

Published online 1 May 2011.

1. Anan, T. *et al.* Human ubiquitin-protein ligase Nedd4: expression, subcellular localization and selective interaction with ubiquitin-conjugating enzymes. *Genes Cells* **3**, 751–763 (1998).
2. Zheng, N., Wang, P., Jeffrey, P. D. & Pavletich, N. P. Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* **102**, 533–539 (2000).
3. Brzovic, P. S. *et al.* Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex. *Proc. Natl Acad. Sci. USA* **100**, 5646–5651 (2003).
4. Huang, A. *et al.* E2-c-Cbl recognition is necessary but not sufficient for ubiquitination activity. *J. Mol. Biol.* **385**, 507–519 (2009).
5. Capili, A. D., Edghill, E. L., Wu, K. & Borden, K. Structure of the C-terminal RING finger from a RING-IBR-RING/TRIAD motif reveals a novel zinc-binding domain distinct from a RING. *J. Mol. Biol.* **340**, 1117–1129 (2004).
6. Shimura, H. *et al.* Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. *Nature Genet.* **25**, 302–305 (2000).

7. Marin, I., Lucas, J. I., Gradilla, A. C. & Ferrus, A. Parkin and relatives: the RBR family of ubiquitin ligases. *Physiol. Genomics* **17**, 253–263 (2004).
8. Tan, N. G. *et al.* Human homologue of ariadne promotes the ubiquitylation of translation initiation factor 4E homologous protein, 4EHP. *FEBS Lett.* **554**, 501–504 (2003).
9. Tokunaga, F. *et al.* Involvement of linear polyubiquitylation of NEMO in NF- κ B activation. *Nature Cell Biol.* **11**, 123–132 (2009).
10. Whitcomb, E. A., Dudek, E. J., Liu, Q. & Taylor, A. Novel control of S phase of the cell cycle by ubiquitin-conjugating enzyme H7. *Mol. Biol. Cell* **20**, 1–9 (2009).
11. Franssen, K. *et al.* Analysis of SNPs with an effect on gene expression identifies UBE2L3 and BCL3 as potential new risk genes for Crohn's disease. *Hum. Mol. Genet.* **19**, 3482–3488 (2010).
12. Pickart, C. M. & Rose, I. A. Functional heterogeneity of ubiquitin carrier proteins. *J. Biol. Chem.* **260**, 1573–1581 (1985).
13. Wang, X. *et al.* Ubiquitination of serine, threonine, or lysine residues on the cytoplasmic tail can induce ERAD of MHC-1 by viral E3 ligase mK3. *J. Cell Biol.* **177**, 613–624 (2007).
14. Cadwell, K. & Coscoy, L. Ubiquitination of nonlysine residues by a viral E3 ubiquitin ligase. *Science* **309**, 127–130 (2005).
15. Christensen, D. E., Brzovic, P. S. & Kleit, R. E. E2-BRCA1 RING interactions dictate synthesis of mono- or specific polyubiquitin chain linkages. *Nature Struct. Mol. Biol.* **14**, 941–948 (2007).
16. Rodrigo-Brenni, M. C. & Morgan, D. O. Sequential E2s drive polyubiquitin chain assembly on APC targets. *Cell* **130**, 127–139 (2007).
17. Jones, D., Crowe, E., Stevens, T. A. & Candido, E. P. Functional and phylogenetic analysis of the ubiquitylation system in *Caenorhabditis elegans*: ubiquitin-conjugating enzymes, ubiquitin-activating enzymes and ubiquitin-like proteins. *Genome Biology* **3**, RESEARCH0002 (2002).
18. Yunus, A. A. & Lima, C. D. Lysine activation and functional analysis of E2-mediated conjugation in the SUMO pathway. *Nature Struct. Mol. Biol.* **13**, 491–499 (2006).
19. Wu, P. Y. *et al.* A conserved catalytic residue in the ubiquitin-conjugating enzyme family. *EMBO J.* **22**, 5241–5250 (2003).
20. Sakata, E. *et al.* Crystal structure of the UbcH5b~ubiquitin intermediate: insight into the formation of the self-assembled E2~Ub conjugates. *Structure* **18**, 138–147 (2010).
21. Ozkan, E., Yu, H. & Deisenhofer, J. Mechanistic insight into the allosteric activation of a ubiquitin-conjugating enzyme by RING-type ubiquitin ligases. *Proc. Natl Acad. Sci. USA* **102**, 18890–18895 (2005).
22. Kamadurai, H. B. *et al.* Insights into ubiquitin transfer cascades from a structure of UbcH5b~Ubiquitin-HECT^{NEDD4L} complex. *Mol. Cell* **36**, 1095–1102 (2009).
23. Ardley, H. C., Tan, N. G., Rose, S. A., Markhan, A. F. & Robinson, P. A. Features of the parkin/ariadne-like ubiquitin ligase, HHARI, that regulates its interaction with the ubiquitin-conjugating enzyme UbcH7. *J. Biol. Chem.* **276**, 19640–19647 (2001).
24. Eddins, M. J., Carlie, C. M., Gomez, K. M., Pickart, C. M. & Wolberger, C. Mms2-Ubc13 covalently bound to ubiquitin reveals the structural basis of linkage-specific polyubiquitin chain formation. *Nature Struct. Mol. Biol.* **13**, 915–920 (2006).
25. Fallon, L. *et al.* A regulated interaction with the UIM protein Eps15 implicates parkin in EGF receptor trafficking and PI(3)K-Akt signalling. *Nature Cell Biol.* **8**, 834–842 (2006).
26. Joch, M. *et al.* Parkin-mediated monoubiquitination of the PDZ protein PICK1 regulates the activity of acid-sensing ion channels. *Mol. Biol. Cell* **18**, 3105–3118 (2007).
27. Chen, D. *et al.* Parkin mono-ubiquitinates Bcl-2 and regulates autophagy. *J. Biol. Chem.* **285**, 38214–38223 (2010).
28. Maruyama, M. *et al.* Novel mutations, pseudo-dominant inheritance, and possible familial affects in patients with autosomal recessive juvenile parkinsonism. *Ann. Neurol.* **48**, 245–250 (2000).
29. Diao, J., Zhang, Y., Huibregtse, J. M., Zhou, D. & Chen, J. Crystal structure of SopA, a *Salmonella* effector protein mimicking a eukaryotic ubiquitin ligase. *Nature Struct. Mol. Biol.* **15**, 65–70 (2008).
30. Dastur, A., Beaudenon, S., Kelley, M., Krug, R. M. & Huibregtse, M. Herc5, an interferon-induced HECT E3 enzyme, is required for conjugation of ISG15 in human cells. *J. Biol. Chem.* **281**, 4334–4338 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Stoll, M. Schwartz and N. Zheng for critical reading of the manuscript and discussions and V. Vittal, K. Stoll and P. Littlefield for technical help. We thank K. Borden, E. Fon, B. Schulman, L. Boyd, J. Chen and A. Merz for expression vectors for HHARI, parkin, E6-AP, Ubc18, human E1 and GST pull-down reagents, respectively. We acknowledge support from the National Institute of General Medical Sciences in the form of 5R01 GM08055 (R.E.K.) and T32 GM07270 (D.M.W.), and the National Science Foundation in the form of MCB0615632 (R.E.K.)

Author Contributions D.M.W. performed all experiments. P.S.B. and A.L. contributed to the experimental design for Fig. 2 and Supplementary Fig. 5. D.M.W. and R.E.K. designed the overall study and wrote the manuscript with P.S.B.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.E.K. (kleit@u.washington.edu).

METHODS

Multiple sequence alignments. Multiple sequence alignments were performed using Clustal W³¹ with manual sequence adjustments based on E2 structures.

Plasmids, protein expression and purification. Plasmid constructs, protein expression, and purification of wheat E1, Ub, UBC13, UBCH5C, UBE2K, UBCH7, Flag-BRCA1 (residues 1–304)/BARD1 (residues 26–327), Flag-BRCA1 (residues 1–112)/BARD1 (residues 26–140) were described previously^{3,15}. Point mutations were introduced using site-directed mutagenesis (Stratagene) and confirmed by DNA sequencing. PFastBac-His-human E1 was expressed in Hi5 cells and purified by Ni²⁺-affinity chromatography, according to the manufacturer's instructions (Sigma), followed by gel filtration using Superdex 200 resin (GE Health Care Life Sciences). E6-AP (residues 495–852) was expressed and purified as described previously³². Ubc18 was sub-cloned into pHis vector³³ in-frame with the N-terminal His-tag, and UBCH7 was subcloned into pet24a vector in-frame with a His-T7 N-terminal tag. His-T7-UBCH7, His-Ubc18, His-UBCH5C(N77S), HA- and T7-tagged Ub were purified by Ni²⁺ affinity chromatography followed by gel-filtration using Superdex 75 resin (GE Healthcare Life Sciences). Constructs of pGEX-4T parkin (rat) and HHARI were expressed in BL21 *Escherichia coli* (Invitrogen) in Luria Broth supplemented with 0.2 mM ZnCl₂ and were purified using GSTrap FF columns (GE Healthcare and Life Sciences) eluted with 10 mM reduced glutathione. Glutathione was removed by dialysis against 50 mM Tris, 200 mM NaCl, 1 mM DTT, pH 7.6 (parkin) and pH 8.0 (HHARI). HHARI_{R2} was subcloned into pGEX-4T in-frame with the N-terminal GST-tag. The GST-tag on HHARI_{R1-IBR-R2} and HHARI_{R2} was removed by thrombin cleavage for NMR and cysteine modification experiments as well as to repeat activity assays shown in Supplementary Fig. 14.

GST pull-down assays. One-hundred microlitre binding reactions contained GST-HHARI (5 μ M) with T7-UBCH7 (5 μ M) and 50 μ l of glutathione sepharose B resin (GE Healthcare) in the binding buffer: 50 mM Tris, 150 mM NaCl, 0.5% Triton X-100, 0.5 mM dithiothreitol (DTT) pH 7.5. Binding reactions were incubated for 3 h at 4 °C, and resin was washed 5 times with 1.5 ml of binding buffer before proteins were eluted with 80 μ l of reduced SDS-PAGE loading buffer. Reaction products were resolved on a 15% SDS-PAGE gel and transferred onto polyvinylidene fluoride membranes (Bio-Rad). The membranes were probed simultaneously with rabbit antibody to GST (Affinity BioReagents) and mouse antibody to T7 (Novagen) followed by goat anti-mouse and goat anti-rabbit secondary antibody conjugated to Alexa Fluor 680 (Molecular probes) and IRdye 800 (Rockland Immunochemicals), respectively. Blotted proteins were detected using an Odyssey infrared imaging system (Licor).

E3 auto-ubiquitination activity assays. One-hundred microlitre reaction mixtures for BRCA1 auto-ubiquitination contained 2 μ M His-Flag BRCA1 (residues 1–304)/BARD1 (residues 26–327), 2 μ M UBCH5C, 20 μ M T7-Ub, 0.5 μ M wheat E1 and 10 mM MgCl₂. Reactions were initiated at 37 °C by adding 10 mM ATP and samples were quenched at the indicated time points by boiling in SDS sample buffer that contained β ME. Ubiquitination products were visualized by western blot, probing for T7 (Ub) and Flag (BRCA1) epitopes simultaneously. Parkin and HHARI ubiquitination assays were performed similarly except products were visualized by probing for the GST tag on parkin or HHARI or the HA epitope on HA-Ub (mouse primary from Covance). HHARI assays with UBCH7, UBCH5C and UBCH5C(N77S/D87K) mutants were performed at higher concentrations (15 μ M E2/E3, 50 μ M Ub) and visualized by Coomassie staining on a 15% SDS-PAGE gel. One-hundred microlitre reaction mixtures for E6-AP thioester formation assays included 15 μ M E2, 15 μ M E6-AP, 30 μ M Ub and 10 mM MgCl₂. Reactions were initiated at 37 °C by the addition of 10 mM ATP and gel samples were taken in parallel at the indicated time points in loading buffer that lacked or contained the reducing agent β ME.

HHARI thioester detection. Reactions containing 20 μ M UBCH7, 20 μ M HA-Ub, 0.5 μ M E1, 10 mM MgCl₂/ATP were incubated at 37 °C for 30 min to form UBCH7-HA-Ub and chilled on ice. Reactions with HHARI were initiated by diluting 2 μ l of charged E2 with 18 μ l of 2 μ M HHARI_{R1-IBR-R2} on ice and incubating

for the indicated times. Reactions were quenched by the addition of SDS-loading buffer that either contained or lacked the reducing agent β ME. HHARI thioesters were visualized by western blot probing for the HA epitope on Ub.

Nucleophile reactivity assays. Reaction mixtures for amino acid reactivity assays contained 20 μ M E2, 20 μ M Ub, 0.5 μ M E1 and 10 mM MgCl₂/ATP in 25 mM NaPi, 150 mM NaCl, pH 7.0 buffer. Amino acids were purchased from Sigma except cysteine (Nutritional Biochemicals Corporation). Five-hundred millimolar stock solutions of L-lysine monohydrochloride, L-arginine monohydrochloride, L-cysteine, L-serine and L-threonine were prepared in reaction buffer and pH was checked by pH paper to be \sim 7. E2s were charged for 20–30 min at 37 °C before being mixed with 50 mM of cysteine, arginine, lysine, serine, threonine or buffer and incubated for 15–20 min at 37 °C. Samples were quenched in non-reducing loading-buffer and visualized by Coomassie stained SDS-PAGE. Reactivity time-courses with lysine and cysteine were performed similarly except that samples were taken at several time points during the reaction.

In reactions containing E3, 20 μ M E3 (His-Flag-BRCA1 residues 1–112 (wild type or I26A)/BARD1 residues 26–140 or E6-AP(C820A) residues 495–852 or GST-HHARI_{R1-IBR-R2}, or GST-HHARI_{R2}) was added and mixed to the pre-charged E2 just before incubation with amino acids. UBCH5C was precharged with Ub where all lysines were mutated to arginines (K0) in reactions with BRCA1 to prevent transfer of Ub. Reactivity reactions visualized by western blot were performed similarly except concentrations were 10 μ M E2 or E3, and 5 μ M HA-Ub.

HHARI oxyster detection. Ubiquitination reactions contained 15 μ M HHARI_{R1-IBR-R2}(C357S), 150 μ M HA-Ub, 1.5 μ M E1, 10 mM MgCl₂ and 15 μ M UBCH7. Reactions were initiated by the addition of 10 mM ATP, and incubated at 37 °C for 30 min before being quenched in reduced SDS-loading buffer. Reactions were then incubated for 20 min at 37 °C with 0.14 N NaOH before being boiled and loaded on a 15% SDS-PAGE gel. Reaction products were visualized by western blot, simultaneously blotting for the HA (Ub) and GST (HHARI) epitopes. For controls, parallel ubiquitination reactions with UBC13 and UBC13(C86S) were performed. UBC13 readily auto-ubiquitinates itself (via an isopeptide) and the UBC13 mutant C86S forms an oxyster-linked Ub conjugate²⁴.

NMR. For the production of ¹⁵N-labelled proteins, bacteria were grown in minimal MOPS medium supplemented with [¹⁵N] ammonium chloride (Cambridge Isotope Labs). NMR data were collected on a Bruker DMX 500 MHz spectrometer. Samples of ¹⁵N-His-Flag-BRCA1(residues 1–112)/BARD1 (residues 26–140) and UBCH5C mutants were prepared as described previously¹⁵. Samples of HHARI_{R2} were prepared as reported previously⁵. Spectra were processed using NMRPipe³⁴/NMRDraw³⁵.

Cysteine modification of HHARI_{R2}. One-hundred microlitre cysteine modification reactions contained 100 μ M wild-type HHARI_{R2} and HHARI_{R2}(C357A) and 500 μ M 4-(2-iodoacetamido)-TEMPO (Sigma). Stock solutions of 4-(2-iodoacetamido)-TEMPO were prepared at 60 mM in DMSO. Cysteine modification reactions were incubated overnight at 4 °C. Samples for MALDI-TOF were diluted 1:10 in MALDI matrix (saturated sinapinic acid (Sigma) in 40% acetonitrile, 0.1% TFA) and masses were quantified by MALDI-TOF spectrometry on a Bruker AutoFlex II spectrometer, using insulin and apomyoglobin as standards.

31. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
32. Huang, L. *et al.* Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *Science* **286**, 1321–1326 (1999).
33. Sheffield, P., Garrard, S. & Derewenda, Z. Overcoming expression and purification problems of RhoGDI using a family of “parallel” expression vectors. *Protein Expr. Purif.* **15**, 34–39 (1999).
34. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
35. Johnson, B. A. & Blevins, R. A. NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).

Substrate-modulated gating dynamics in a Na^+ -coupled neurotransmitter transporter homologue

Yongfang Zhao^{1,2,4*}, Daniel S. Terry^{5*}, Lei Shi^{5,6*}, Matthias Quick^{1,2,4}, Harel Weinstein^{5,6}, Scott C. Blanchard⁵ & Jonathan A. Javitch^{1,2,3,4}

Neurotransmitter/ Na^+ symporters (NSSs) terminate neuronal signalling by recapturing neurotransmitter released into the synapse in a co-transport (symport) mechanism driven by the Na^+ electrochemical gradient^{1–6}. NSSs for dopamine, noradrenaline and serotonin are targeted by the psychostimulants cocaine and amphetamine¹, as well as by antidepressants⁷. The crystal structure of LeuT, a prokaryotic NSS homologue, revealed an occluded conformation in which a leucine (Leu) and two Na^+ are bound deep within the protein⁸. This structure has been the basis for extensive structural and computational exploration of the functional mechanisms of proteins with a LeuT-like fold^{9–22}. Subsequently, an ‘outward-open’ conformation was determined in the presence of the inhibitor tryptophan²³, and the Na^+ -dependent formation of a dynamic outward-facing intermediate was identified using electron paramagnetic resonance spectroscopy²⁴. In addition, single-molecule fluorescence resonance energy transfer imaging has been used to reveal reversible transitions to an inward-open LeuT conformation, which involve the movement of transmembrane helix TM1a away from the transmembrane helical bundle²². We investigated how substrate binding is coupled to structural transitions in LeuT during Na^+ -coupled transport. Here we report a process whereby substrate binding from the extracellular side of LeuT facilitates intracellular gate opening and substrate release at the intracellular face of the protein. In the presence of alanine, a substrate that is transported ~10-fold faster than leucine^{15,25}, we observed alanine-induced dynamics in the intracellular gate region of LeuT that directly correlate with transport efficiency. Collectively, our data reveal functionally relevant and previously hidden aspects of the NSS transport mechanism that emphasize the functional importance of a second substrate (S2) binding site within the extracellular vestibule^{15,20}. Substrate binding in this S2 site appears to act cooperatively with the primary substrate (S1) binding site to control intracellular gating more than 30 Å away, in a manner that allows the Na^+ gradient to power the transport mechanism.

The experiments were performed on LeuT engineered to contain a 15-amino-acid, carboxy-terminal biotinylation domain²⁶ and site-specifically labelled with the fluorophores Cy3 and Cy5 maleimide at residue position 7, after replacing the native His residue with Cys (H7C) in the amino-terminal loop close to TM1, and at position 86 (R86C) in intracellular loop (IL) 1 (Methods). Direct observations of conformational processes within the intracellular gate region of LeuT (Supplementary Fig. 1) were made using a wide field imaging strategy employing prism-based total internal reflection (Methods, Fig. 1a). As described²², fluorescence resonance energy transfer (FRET) imaging of

LeuT revealed two readily distinguished states (FRET efficiency ~0.51 and ~0.75) in the presence of 200 mM K^+ and the nominal absence of Na^+ (Fig. 1b), consistent with the existence of two distinct conformations of the intracellular gate that differ by ~13 Å in the distance separating the fluorophore pair.

In experiments imaging LeuT dynamics with increasing Na^+ concentrations, Hidden Markov Modelling revealed that the distribution of low- and high-FRET conformations of LeuT was altered by Na^+ with an effector concentration for half-maximum response (EC_{50}) of 10.9 mM (Fig. 1b, c), consistent with the EC_{50} for Na^+ -dependent stimulation of substrate binding and transport¹⁵. Na^+ decreased the overall frequency of transitions (Fig. 1d, e) through the preferential stabilization (~7-fold) of the inward-closed state. During the direct imaging of individual LeuT molecules (Fig. 1f), slow, spontaneous transitions between open and closed states, initially observed in 200 mM K^+ , were dramatically decreased on exchange into Na^+ -containing buffer, leading to the preferential stabilization of the inward-closed state.

Reasoning that substrate-induced intracellular gating might be observed best under conditions mimicking the relatively low intracellular Na^+ , we performed experiments at Na^+ concentrations sufficient for Leu binding but below the EC_{50} of Na^+ . However, even at 2 mM Na^+ , Leu shifted the population towards the closed intracellular gate conformation (Supplementary Fig. 2a, b) through a ~3.5-fold stabilization of this state (Supplementary Fig. 2c). These effects, which result in a global decrease in transition frequency (Supplementary Fig. 2d), were recapitulated at the level of individual LeuT molecules (Supplementary Fig. 2e). Thus, while unambiguously demonstrating binding of both Na^+ and Leu to LeuT, these results corroborate our earlier finding that Leu binding has the net effect of diminishing the likelihood of intracellular gate opening. One possible explanation for these observations is that Leu’s high affinity for the transporter¹⁵ makes it a poor substrate for transport, which in our measurements is manifested in the greatly extended lifetime of the closed state. To test this hypothesis, intracellular gate dynamics were assessed in the presence of the more efficiently transported substrate Ala.

In stark contrast to Leu, under otherwise identical conditions, increasing Ala concentrations did not shift the FRET distribution towards the closed state (Fig. 2a, b). Instead, a strong, Ala-concentration-dependent enhancement of transition rates was observed. In 2 mM Na^+ , Ala enhanced the transition rates between inward-open and inward-closed states by as much as ~4-fold (Fig. 2c, d). This result was directly confirmed at the scale of individual molecules on exchange into Ala-containing buffer (Fig. 2e). Similar enhancements in transition frequency were also

¹Center for Molecular Recognition, Columbia University College of Physicians and Surgeons, 630 West 168th Street, New York, New York 10032, USA. ²Department of Psychiatry, Columbia University College of Physicians and Surgeons, 630 West 168th Street, New York, New York 10032, USA. ³Department of Pharmacology, Columbia University College of Physicians and Surgeons, 630 West 168th Street, New York, New York 10032, USA. ⁴Division of Molecular Therapeutics, New York State Psychiatric Institute, 1051 Riverside Drive, New York, New York 10032, USA. ⁵Department of Physiology and Biophysics, Weill Medical College of Cornell University, 1300 York Avenue, New York, New York 10021, USA. ⁶HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, Cornell University, 1300 York Avenue, New York, New York 10021, USA.

*These authors contributed equally to this work.

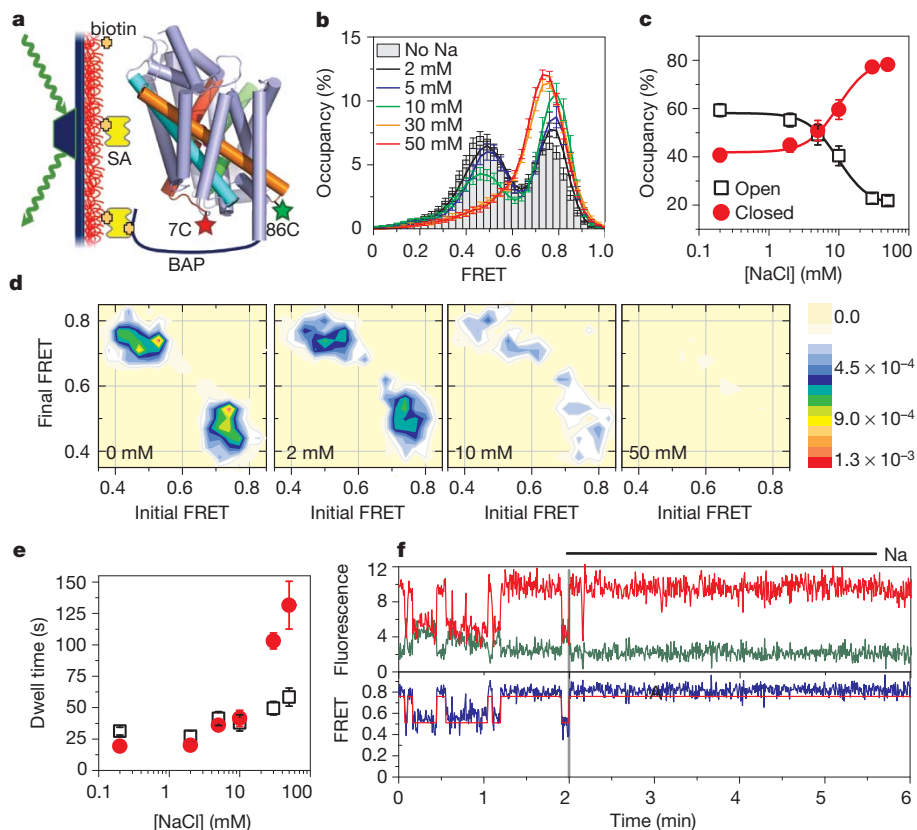


Figure 1 | Effect of Na⁺ on LeuT dynamics. **a**, Experimental set-up: H7C/R86C-LeuT labelled with Cy3 and Cy5 (stars) was immobilized via a biotin acceptor peptide (BAP) on a passivated glass surface and illuminated using total internal reflection. FRET traces (>110 per condition) were collected with varying concentrations of Na⁺ (160-ms time resolution for all, except 30–50 mM with 400 ms). **b**, Histograms of FRET traces, filtered to remove fluorophore dark states. **c**, Fraction of time in the lower-FRET open state (black

open squares) and the high-FRET closed state (red filled circles). **d**, Transition density plot: average FRET values before (x axis) and after (y axis) each transition were plotted as a two-dimensional chart in transitions per second (scale at right; Na⁺ concentrations are indicated). **e**, Average dwell times in each state.

f, Representative traces (donor in green, acceptor in red, FRET in blue, and predicted state sequence (idealization) in red), where the solution was exchanged at 2 min from K⁺ to Na⁺ (200 mM). Error bars, s.d. of ≥ 100 bootstrap samples.

observed for H7C/T515C-LeuT (Supplementary Fig. 3). In accordance with such effects, which required both Na⁺ and Ala, the lifetimes of the inward-open or inward-closed FRET states were not significantly affected by Ala alone (in the nominal absence of Na⁺); at 250 μ M Ala, the transition frequency increased in a Na⁺-concentration-dependent fashion (Supplementary Fig. 4).

Using transition state theory (Methods), we found that the intracellular open and closed FRET states of LeuT were separated by a large activation barrier ($\Delta G^\ddagger \approx 80$ kJ mol⁻¹). Ala does not alter the relative occupancies of open and closed states, but instead lowers the activation barrier for both open-to-closed and closed-to-open transitions by approximately 3 kJ mol⁻¹ (about the energy of a hydrogen bond). By contrast, Leu raised the activation barrier for the closed-to-open transition by as much as 4 kJ mol⁻¹, apparently through ground-state stabilization of the closed state.

Hypothesizing that the observed dynamics reflect Ala's acceleration of the opening-closing cycles of the intracellular gate required for the transport mechanism, we performed experiments in the presence of the transport inhibitor clomipramine (CMI), a tricyclic antidepressant that is known to bind in an extracellular vestibule above the Na⁺ and S1 binding sites^{25,27,28}. Many of the residues shown to interact with antidepressants in these structures are also part of the S2 site^{25,27}. As substrate binding in the S2 site is thought to allosterically trigger intracellular release of Na⁺ and substrate from the S1 site¹⁵ (also see Supplementary Fig. 1), CMI should block Ala-induced intracellular gating dynamics. Indeed, in the presence of both Na⁺ and Ala, CMI essentially eliminated intracellular gate opening, stabilizing LeuT in a high-FRET, inward-closed conformation (Supplementary Fig. 5a–c).

This observation is consistent with CMI competitively blocking substrate binding to the S2 site¹⁵, thereby preventing Ala-induced opening and closing of the intracellular gate, and inhibiting transport. This result was again confirmed by direct imaging of individual LeuT molecules in Na⁺ and Ala-containing buffer on addition of CMI (Fig. 2f). The detergent n-octyl- β -D-glucopyranoside also inhibited intracellular gating dynamics (Supplementary Fig. 5a–c), consistent with its capacity to disrupt the Na⁺-coupled transport mechanism²⁰ by competing with substrate binding to the S2 site^{20,23}.

To probe whether Ala binding to the S1 and/or S2 site(s) was responsible for lowering the activation barrier for intracellular gating dynamics, single-molecule FRET experiments were performed in the background of either an F253A or L400S mutation within the S1 or S2 site, respectively (Fig. 3a, Supplementary Fig. 1). These mutations disrupt substrate binding to LeuT, decreasing the stoichiometry of substrate:LeuT binding under saturating conditions from 2:1 in wild-type LeuT, to 1:1 in both mutants (Fig. 4a). Mutation of F253 blocks substrate binding to the S1 site and also abrogates transport (Fig. 4a, b; Supplementary Fig. 6), while having little or no effect on Na⁺ binding (Supplementary Table 1). Despite evidence that Ala bound to the S2 site in the context of the F253A mutation (Fig. 4a), Ala failed to increase intracellular gating dynamics of the mutant protein (Fig. 4c). Similarly, despite evidence of Ala binding to the S1 site (Fig. 4a), no increase in intracellular gating dynamics was observed when the S2 site was disrupted by the L400S mutation (Fig. 4c). These findings support the notion that substrate occupancy in the S2 site is critical for the allosteric mechanism that controls intracellular gate opening and the release of substrate from the S1

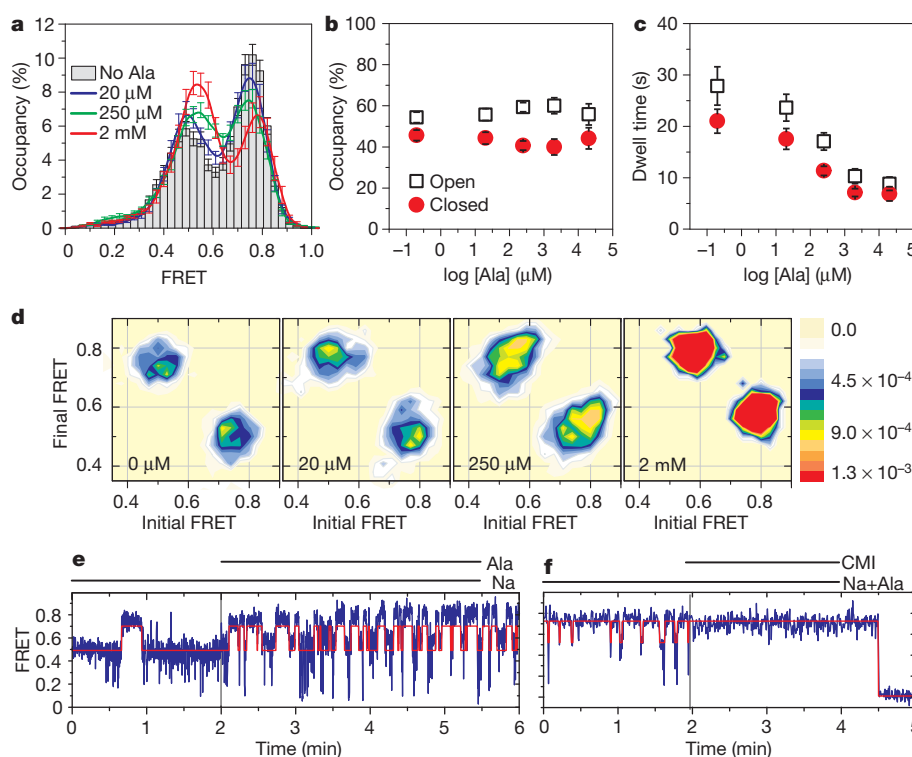


Figure 2 | Effect of alanine on LeuT dynamics. Single-molecule FRET traces (>90 per condition) were collected at 160-ms time resolution with 2 mM Na^+ and varying concentrations of Ala. **a**, Histograms of FRET data from each condition. Hidden Markov Modelling analysis revealed the fraction of time (**b**) and average dwell times (**c**) in the lower-FRET open state (black open squares) and the high-FRET closed state (red filled circles). **d**, Transition

density plots as in Fig. 1d (Ala concentrations are indicated). **e**, **f**, Representative FRET traces (blue) with idealization (red) from experiments where solution was exchanged at 2 min: **e**, 2 mM Na^+ , adding 250 μM Ala; **f**, 2 mM Na^+ and 250 μM Ala, adding the inhibitor clomipramine (CMI; 0.5 mM). Error bars, s.d. of ≥ 100 bootstrap samples.

site¹⁵, and demonstrate that substrate binding to both the S1 and S2 sites is necessary to trigger intracellular gating.

In order to probe whether Ala binding to the S1 and S2 sites is also sufficient to promote intracellular gating and transport, experiments were performed in the presence of Li^+ in place of Na^+ . In the presence of saturating Li^+ concentrations (>150 mM), we found that Ala binds LeuT with a 2:1 stoichiometry consistent with both S1 and S2 site occupancy (Fig. 4a). Li^+ , like Na^+ , stabilized the inward-closed state (Supplementary Fig. 7), but, in the presence of Li^+ , Ala failed to accelerate intracellular gating dynamics and no substrate transport was observed (Fig. 4c). Instead, the inward-closed conformation of LeuT

was modestly stabilized in the presence of Ala (~ 2 -fold reduction in the rate of gate opening, $k_{\text{closed-open}}$) (Fig. 4c). These data demonstrate that Ala binding to the S1 and S2 sites in the presence of Li^+ does not lower the activation barrier to intracellular gating as observed in the presence of Na^+ .

Prompted by these experimental observations, computational studies were performed to investigate how both Na^+ and Li^+ can support substrate binding to LeuT, whereas only Na^+ leads to substrate-induced dynamics of the intracellular gate and to transport. These studies also served to identify both local changes produced in the region of the ion binding sites and critical elements in the allosteric pathway linking the

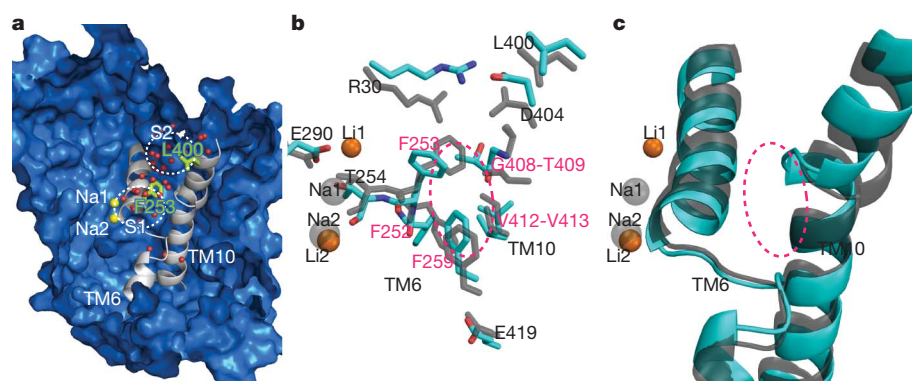


Figure 3 | The configuration of TM6–TM10 interactions induced by Na^+ binding cannot be matched by Li^+ binding. **a**, Representative snapshot taken from the Na-only simulation, showing water molecules (red spheres) occupying the S1 and S2 sites (white dotted ellipses). Residues L400 in the S2 site and F253 in the S1 site, which were mutated to affect substrate binding, are shown as light green sticks. **b**, The different effects that Li^+ and Na^+ binding

have on the interacting residues of TM6 and TM10. The TM6/TM10 interface is indicated by the dashed ellipse in magenta. **c**, The bulge around G408 in TM10, which is present only when Na^+ is bound but not when Li^+ replaces it. In **b** and **c**, side chains and backbones coloured according to atom types are from the Li-only conformation, while those from the Na-only conformation are rendered in grey.

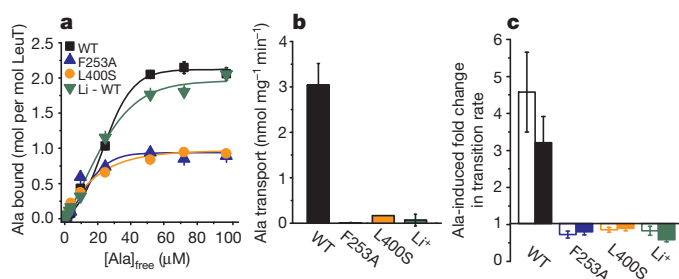


Figure 4 | Effect of S1 and S2 site mutations and of Li^+ on activity and dynamics. **a**, Binding of ^3H -Ala in buffer containing 50 mM Na^+ was measured for wild-type (WT, black squares), F253A (blue triangles) and L400S (orange circles) LeuT, and for wild-type LeuT with 150 mM Li^+ (green inverted triangles). **b**, ^3H -Ala uptake with 100 mM Na^+ was measured for WT (black), F253A (blue) and L400S (orange) LeuT or with 150 mM Li^+ for WT LeuT (green). Error bars in **a** and **b**, s.e.m. of triplicate determinations. **c**, The fold change in the rate of transitioning from the open state to the closed state (open bars) and from the closed state to the open state (filled bars) induced by 250 μM Ala in 10 mM Na^+ or for WT in 40 mM Li^+ (>280 traces and 800 transitions per condition). Error bars in **c**, s.d. of ≥ 100 bootstrap samples.

substrate binding sites and the intracellular gate region. Comparative analysis of separate molecular dynamics (MD) simulations of LeuT, performed with either Na^+ or Li^+ occupying the established Na^+ binding sites and in the absence of amino acid substrate (termed Na-only²⁴ and Li-only, respectively) revealed significant differences in TM-TM interactions (Fig. 3b, c), which are described in detail in Supplementary Information. The Na1/Li1 binding site and its neighbouring interaction network, which are crucial for the proper propagation of the allosteric effects from the S2 to S1 site (see Supplementary Information for details) and onward to the intracellular side to open the transport pathway, are sensitive to the unique combination of the ionic radius of the Na^+ cation and the charge redistribution it causes. The structural consequences of the ion-specific effects appear to be propagated through the cluster of aromatic residues at the heart of the S1 binding site, and result in different configurations of the bulge in the middle of TM10 (Fig. 3, Supplementary Fig. 8).

The positions of the structural elements involved in this propagation mechanism make them critical for transmitting conformational changes deeper into the TM bundle towards the intracellular end of the transporter (Fig. 3b, c). Such changes include local alterations in the vicinity of E419, a residue known from the crystal structure to interact with E62 in TM2, with the backbone of the unwound portion of TM6 (proximal to F259 of the aromatic cluster and the S1 binding site), and two water molecules²⁹. Reconfiguration of this region, including residue T418, on simulated inward movement of the substrate¹⁵ was previously shown to enable the penetration of water from the intracellular side of LeuT as a result of an opening at IL1¹⁵. The resulting dissociation of IL1 from interactions with R5 and D369 and the destabilization of the network of intracellular interactions detected in the simulations (Supplementary Fig. 9) is associated with the observed outward movement of TM1a²² that is essential for the simulated release of substrate to the intracellular side.

Owing to the different effects of Li^+ and Na^+ , Ala binding in both the S1 and S2 sites in the presence of Li^+ would not engender the ordered series of local conformational rearrangements expected in the presence of Na^+ . These rearrangements originate in the S2 site and need to be propagated as described above through changes in the Na1 and S1 sites to enable water penetration from the cytoplasmic side of LeuT and the outward movement of TM1a. Their absence when Li^+ substitutes for Na^+ would explain why substrate-induced acceleration of gating dynamics was not observed experimentally.

Na^+ binding, which stabilizes the inward-closed state, does not hasten gate closure but, instead, slightly stabilizes the inward-open state as well, by raising the energy barrier to the conformational transition. In contrast, Ala binding to LeuT in the presence of Na^+ shortens

not only the inward-closed, but also the inward-open, lifetime (Fig. 2). Thus, bound Ala facilitates both the opening of the intracellular gate and its subsequent closure by reducing the activation barrier for such conformational transitions. One possible explanation for this observation is that binding of substrate in the S2 site triggers the opening of the intracellular gate and release of the S1 substrate to the cytoplasm. In the absence of S1 substrate and bound Na^+ , substrate in the S2 site may then facilitate intracellular gate closure. It is tempting to speculate that the S2 substrate, in the presence of extracellular Na^+ , may move to the S1 site with high efficiency owing to its very high local concentration, thereby facilitating a subsequent transport cycle.

Collectively, our findings support the notion that the observed movements of TM1a and its environment are associated with LeuT intracellular gating²² in a manner that is directly linked to the Na^+ -driven transport mechanism. Thus, results obtained with the slowly transported substrate, Leu, and the relatively rapidly transported substrate, Ala, establish a relationship between the rates of intracellular gating and substrate transport. The role of substrate binding at the S2 site in the process of allostery and molecular recognition is further highlighted by the comparative effects of CMI and Ala binding to this site in the presence of Na^+ . The former stabilizes a closed intracellular gate conformation, whereas the latter substantially lowers the activation barrier to gate opening and thereby allows the energy of the Na^+ gradient to drive the transport mechanism.

After the present manuscript had been submitted, a report³⁰ was published that concluded, on the basis of a variety of binding measurements, that LeuT has only a single high-affinity substrate site. In contrast, our substrate binding measurements clearly show a stoichiometry of 2:1, consistent with high affinity binding to both the S1 and S2 sites²⁰. Half of this binding is lost in the S2-site mutant that also exhibits a loss of substrate-induced single-molecule dynamics and transport (Fig. 4). Although the loss of substrate-induced dynamics and transport in the S2-site mutant could conceivably be explained solely by a long-range allosteric effect of the mutation, all our data to date are most consistent with a two-substrate-site model in which the absence of either S1 or S2 substrate binding results in a profound attenuation of transporter dynamics and function. We are currently uncovering the reasons for the discrepancy between the data of ref. 30 and our own data, and will report our findings in due course.

METHODS SUMMARY

LeuT mutants were expressed in *Escherichia coli*, purified, and labelled on targeted engineered cysteines with Cy3 and Cy5 maleimide. The functional properties of the labelled constructs were determined by measuring Leu binding and Na^+ by scintillation proximity assay, and Ala transport was measured after reconstitution of the protein into proteoliposomes. Purified, labelled protein was immobilized onto a passivated glass surface via a streptavidin-biotin linkage (shown schematically in Fig. 1a). Fluorescence data were acquired using a prism-based total internal reflection (TIR) microscope. FRET efficiency was calculated and analysis of fluorescence and FRET traces was achieved using semi-automated analysis software developed for this application. The single-molecule traces were analysed for LeuT in the presence and absence of the substrates Na^+ , Leu and Ala, and on addition of the transport inhibitors CMI and n-octyl- β -D-glucopyranoside, and in response to mutations of the S1 and S2 binding sites. MD simulations were carried out with the protein immersed in an all-atom model of the membrane, solvated with water molecules, ions and ligands. Long equilibrations (totalling >2 μs) were run to assess conformational changes, with more than one MD trajectory collected for every configuration mentioned.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 August 2010; accepted 25 February 2011.

Published online 24 April 2011; corrected 2 June 2011 (see full-text HTML version for details).

- Amara, S. G. & Sonders, M. S. Neurotransmitter transporters as molecular targets for addictive drugs. *Drug Alcohol Depend.* **51**, 87–96 (1998).
- Rudnick, G. *Mechanisms of Biogenic Amine Neurotransmitter Transporters* 2nd edn (Humana, 2002).

3. Sonders, M. S., Quick, M. & Javitch, J. A. How did the neurotransmitter cross the bilayer? A closer view. *Curr. Opin. Neurobiol.* **15**, 296–304 (2005).
4. Gu, H., Wall, S. C. & Rudnick, G. Stable expression of biogenic amine transporters reveals differences in inhibitor sensitivity, kinetics, and ion dependence. *J. Biol. Chem.* **269**, 7124–7130 (1994).
5. Torres, G. E., Gainetdinov, R. R. & Caron, M. G. Plasma membrane monoamine transporters: structure, regulation and function. *Nature Rev. Neurosci.* **4**, 13–25 (2003).
6. Krause, S. & Schwarz, W. Identification and selective inhibition of the channel mode of the neuronal GABA transporter 1. *Mol. Pharmacol.* **68**, 1728–1735 (2005).
7. Iversen, L. Neurotransmitter transporters and their impact on the development of psychopharmacology. *Br. J. Pharmacol.* **147** (Suppl 1), S82–S88 (2006).
8. Yamashita, A. *et al.* Crystal structure of a bacterial homologue of Na⁺/Cl[−]-dependent neurotransmitter transporters. *Nature* **437**, 215–223 (2005).
9. Beuming, T., Shi, L., Javitch, J. A. & Weinstein, H. A comprehensive structure-based alignment of prokaryotic and eukaryotic neurotransmitter/Na⁺ symporters (NSS) aids in the use of the LeuT structure to probe NSS structure and function. *Mol. Pharmacol.* **70**, 1630–1642 (2006).
10. Quick, M. *et al.* State-dependent conformations of the translocation pathway in the tyrosine transporter Tyt1, a novel neurotransmitter:sodium symporter from *Fusobacterium nucleatum*. *J. Biol. Chem.* **281**, 26444–26454 (2006).
11. Forrest, L. R. *et al.* Mechanism for alternating access in neurotransmitter transporters. *Proc. Natl Acad. Sci. USA* **105**, 10338–10343 (2008).
12. Kniazeff, J. *et al.* An intracellular interaction network regulates conformational transitions in the dopamine transporter. *J. Biol. Chem.* **283**, 17691–17701 (2008).
13. Noskov, S. Y. Molecular mechanism of substrate specificity in the bacterial neutral amino acid transporter LeuT. *Proteins* **73**, 851–861 (2008).
14. Noskov, S. Y. & Roux, B. Control of ion selectivity in LeuT: two Na⁺ binding sites with two different mechanisms. *J. Mol. Biol.* **377**, 804–818 (2008).
15. Shi, L. *et al.* The mechanism of a neurotransmitter:sodium symporter — inward release of Na⁺ and substrate is triggered by substrate in a second binding site. *Mol. Cell* **30**, 667–677 (2008).
16. Singh, S. K. LeuT: a prokaryotic stepping stone on the way to a eukaryotic neurotransmitter transporter structure. *Channels (Austin)* **2**, 380–389 (2008).
17. Crisman, T. J., Qu, S., Kanner, B. I. & Forrest, L. R. Inward-facing conformation of glutamate transporters as revealed by their inverted-topology structural repeats. *Proc. Natl Acad. Sci. USA* **106**, 20752–20757 (2009).
18. Khalili-Araghi, F. *et al.* Molecular dynamics simulations of membrane channels and transporters. *Curr. Opin. Struct. Biol.* **19**, 128–137 (2009).
19. Li, J. & Tajkhorshid, E. Ion-releasing state of a secondary membrane transporter. *Biophys. J.* **97**, L29–L31 (2009).
20. Quick, M. *et al.* Binding of an octylglucoside detergent molecule in the second substrate (S2) site of LeuT establishes an inhibitor-bound conformation. *Proc. Natl Acad. Sci. USA* **106**, 5563–5568 (2009).
21. Shi, L. & Weinstein, H. Conformational rearrangements to the intracellular open states of the LeuT and ApcT transporters are modulated by common mechanisms. *Biophys. J.* **99**, L103–L105 (2010).
22. Zhao, Y. *et al.* Single-molecule dynamics of gating in a neurotransmitter transporter homologue. *Nature* **465**, 188–193 (2010).
23. Singh, S. K., Piscitelli, C. L., Yamashita, A. & Gouaux, E. A competitive inhibitor traps LeuT in an open-to-out conformation. *Science* **322**, 1655–1661 (2008).
24. Claxton, D. P. *et al.* Ion/substrate-dependent conformational dynamics of a bacterial homolog of neurotransmitter:sodium symporters. *Nature Struct. Mol. Biol.* **17**, 822–829 (2010).
25. Singh, S. K., Yamashita, A. & Gouaux, E. Antidepressant binding site in a bacterial homologue of neurotransmitter transporters. *Nature* **448**, 952–956 (2007).
26. Beckett, D., Kovaleva, E. & Schatz, P. J. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* **8**, 921–929 (1999).
27. Zhou, Z. *et al.* LeuT-desipramine structure reveals how antidepressants block neurotransmitter reuptake. *Science* **317**, 1390–1393 (2007).
28. Zhou, Z. *et al.* Antidepressant specificity of serotonin transporter suggested by three LeuT-SSRI structures. *Nature Struct. Mol. Biol.* **16**, 652–657 (2009).
29. Sen, N., Shi, L., Beuming, T., Weinstein, H. & Javitch, J. A. A pincer-like configuration of TM2 in the human dopamine transporter is responsible for indirect effects on cocaine binding. *Neuropharmacology* **49**, 780–790 (2005).
30. Piscitelli, C. L., Krishnamurthy, H. & Gouaux, E. Neurotransmitter/sodium symporter orthologue LeuT has a single high-affinity substrate site. *Nature* **468**, 1129–1132 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Altman for assistance in preparing reagents for single-molecule experiments and F. Carvalho for the preparation of membranes. Molecular graphics were prepared with PyMOL. Computations were performed on Ranger at the Texas Advanced Computing Center (TG-MCB090022) and the David A. Cofrin computational infrastructure of the Institute for Computational Biomedicine at Weill Cornell Medical College. This work was supported in part by National Institutes of Health grants DA17293 and DA022413 (J.A.J.), DA12408 (H.W.), DA023694 (L.S.), the Irma T. Hirsch/Monique Weill-Caulier trusts (S.C.B.) and the Lieber Center for Schizophrenia Research and Treatment. D.S.T. is supported by the Tri-Institutional Training Program in Computational Biology and Medicine.

Author Contributions Y.Z. expressed, purified and labelled the LeuT mutants. M.Q. and Y.Z. performed the functional characterization of the mutants. Y.Z. and D.S.T. designed, carried out and analysed the single-molecule experiments; L.S. and H.W. designed and analysed the computational studies, which were carried out by L.S.; S.C.B. and J.A.J. helped to design the biochemical and single-molecule experiments and, with L.S. and H.W., helped to interpret the data. All the authors contributed to writing and editing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.A.J. (jjaj2@columbia.edu) or S.C.B. (scb2005@med.cornell.edu).

METHODS

Protein expression and purification. LeuT variants were expressed in *E. coli* C41(DE3) as described¹⁵. For functional studies, LeuT variants were expressed from pQO18 or derivatives thereof carrying the indicated mutations²⁰, whereas for single-molecule FRET studies, biotin acceptor peptide-tagged LeuT variants were expressed in pETO18G and its derivatives²². Protein was purified by immobilized metal (Ni^{2+}) affinity chromatography using a Ni^{2+} Sepharose 6 FastFlow column (GE Healthcare)²². For fluorescent labelling of LeuT, Cy3-maleimide and Cy5-maleimide (GE Healthcare) were added at an equimolar ratio (200 μM total) for 1 h while the protein was bound to the Ni^{2+} resin²². Free dye was removed before the elution of LeuT with 300 mM imidazole.

Scintillation proximity-based binding studies. Binding of ^3H -leucine or ^3H -alanine (146 Ci mmol^{-1} and 49.4 Ci mmol^{-1} , respectively; both from Moravek) to purified LeuT-variants was measured with the scintillation proximity assay (SPA) as described¹⁵, with 25 ng of purified protein per assay in buffer composed of 150 mM Tris/MES, pH 7.5/50 mM NaCl/1 mM TCEP/0.1% n-dodecyl- β -D-maltopyranoside or 50 mM Tris/MES, pH 7.5/150 mM LiCl/1 mM TCEP/0.1% n-dodecyl- β -D-maltopyranoside. To determine the molar ratio of Leu (or Ala)-to LeuT, binding samples were incubated with increasing concentrations of radioligand and measured in the SPA c.p.m. mode of the Wallac 1450 MicroBeta counter (Perkin Elmer). The efficiency of detection was calculated with standard curves of known concentrations of ^3H -Leu or ^3H -Ala. The standard curves were used to transform c.p.m. into the amount of bound substrate¹⁵. The amount of LeuT in the SPA assays was determined³¹. SPA-based binding studies using 2 μM [^{22}Na]Cl (1,017 mCi mg^{-1} ; Perkin Elmer) were performed in 150–200 mM Tris/MES, pH 7.5/1 mM TCEP/0.1% n-dodecyl- β -D-maltopyranoside in the presence of 0–50 mM NaCl (equimolar replacement of Tris/MES to maintain a total molarity of 200 mM)¹⁵. All experiments were repeated at least in duplicate with triplicate determination of all individual data points. Kinetic constants (shown \pm the s.e.m. of the fit) were obtained by fitting the data from independent experiments to global fitting in Prism or SigmaPlot.

^3H -Ala transport in proteoliposomes. Proteoliposomes were prepared as described¹⁵. The accumulation of ^3H -Ala (49.4 Ci mmol^{-1} ; Moravek) was measured at 23 °C in assay buffer composed of 150/50 mM Tris/MES (pH 8.5) and 50 mM NaCl/150 mM LiCl. The reaction was quenched by the addition of ice-cold assay buffer without radiotracer and the proteoliposomes were collected on GF-75 glass fibre filters (Advantec) before the determination of the accumulated c.p.m. by liquid scintillation counting.

Single-molecule FRET imaging experiments. Fluorescence experiments were performed using a prism-based total internal reflection fluorescence (TIRF) microscope as previously described^{22,32}. Microfluidic imaging chambers were passivated with a mixture of PEG and biotin-PEG and incubated with 0.8 μM streptavidin (Invitrogen). Cy3/Cy5-labelled, biotinylated LeuT molecules were surface immobilized through biotin-streptavidin interaction. Cy3 fluorophores were excited by the evanescent wave generated by total internal reflection (TIR) of a single-frequency light source (Ventus 532, Laser Quanta). Photons emitted from Cy3 and Cy5 were collected using a 1.2 NA 60 \times water-immersion objective (Nikon) and optical treatments were used to separate Cy3 and Cy5 frequencies onto a cooled, back-thinned EMCCD camera (Cascade 512, Photometrics). Fluorescence data were acquired using Metamorph (Universal Imaging Corporation).

All experiments were performed in buffer containing 50 mM Tris/MES, pH 7.5, 10% glycerol, 0.02% (w/v) DDM, 5 mM 2-mercaptoethanol and 200 mM salt (KCl

or NaCl, as specified). We used an oxygen-scavenging environment (1 unit per ml glucose oxidase, 8 units per ml catalase, 0.1% (v/v) glucose) containing 2 mM cyclooctatetraene in all experiments to minimize photobleaching.

Analysis of single-molecule fluorescence data was performed using custom software written in MATLAB (MathWorks). A subset of the acquired traces was selected for further analysis using the following criteria: (1) single-step donor photobleaching, (2) signal-to-background noise ratio (SNR) ≥ 8 , (3) <4 donor blinking events, (4) non-zero FRET efficiency for at least 60 s. Additional manual trace selection was performed to refine the data, where selected traces were required to have: (1) stable total fluorescence intensity ($I_D + I_A$) and (2) at least one transition between clearly defined FRET states with anti-correlated transitions in donor/acceptor intensity or a single dwell in a clearly-defined FRET state. We found this process to be effective in removing artefacts and spurious noise without introducing significant bias (see Supplementary Discussion and Supplementary Fig. 10).

Kinetic analysis was performed to idealize FRET traces and calculate average dwell times using a three-state model as previously described²². Error bars for transition rate estimates and FRET histograms were calculated as the standard deviation of 100 bootstrap samples of the traces. Error bars for state occupancies were calculated from 1,000 bootstrap samples.

Transition rates were interpreted using transition state theory, where the open and closed states are considered ground states separated by a large ($\Delta G^\ddagger \approx 80 \text{ kJ mol}^{-1}$) activation barrier (the transition state). The energy required to achieve the transition state (and cross the barrier) was calculated as:

$$\Delta G^\ddagger = -RT \ln \left(\frac{hk_{ij}}{k_B T} \right),$$

where R is the gas constant, T is absolute temperature (296 K), h is Planck's constant, k is the measured transition rate, from state i to state j , and k_B is Boltzmann's constant. Changes in the activation barrier energy ($\Delta\Delta G^\ddagger$) were calculated from the difference in forward and reverse rates observed in the absence and presence of substrate.

Molecular dynamics. The Li^+ -only simulation was performed on a system prepared as described²⁴. Briefly, it consisted of more than 77,000 atoms, including the explicit membrane model, solvating water molecules and the various ions and ligands. All the Na^+ ions in the system were replaced with Li^+ . The parameters for Li^+ were from ref. 33. All MD simulations were carried out with the NAMD program under constant temperature (310 K) and constant pressure (1 atm) conditions. Long equilibration runs were performed to allow the system to transition to a new stable conformation. The inward-closed and inward-open conformations described in Supplementary Fig. 9 were based on the simulations described previously²². More than one MD trajectory was collected for every configuration studied. Each individual trajectory was at least 360 ns, and the longest trajectory for each configuration was 720 ns. All the results reinforced the conclusions, and the structural and dynamic insights described in the main text were revealed as the common features and trends of parallel independent MD runs.

- Schaffner, W. & Weissmann, C. A rapid, sensitive, and specific method for the determination of protein in dilute solution. *Anal. Biochem.* **56**, 502–514 (1973).
- Munro, J. B., Altman, R. B., O'Connor, N. & Blanchard, S. C. Identification of two distinct hybrid state intermediates on the ribosome. *Mol. Cell* **25**, 505–517 (2007).
- Caplan, D. A., Subbotina, J. O. & Noskov, S. Y. Molecular mechanism of ion-ion and ion-substrate coupling in the Na^+ -dependent leucine transporter LeuT. *Biophys. J.* **95**, 4613–4621 (2008).

ERRATUM

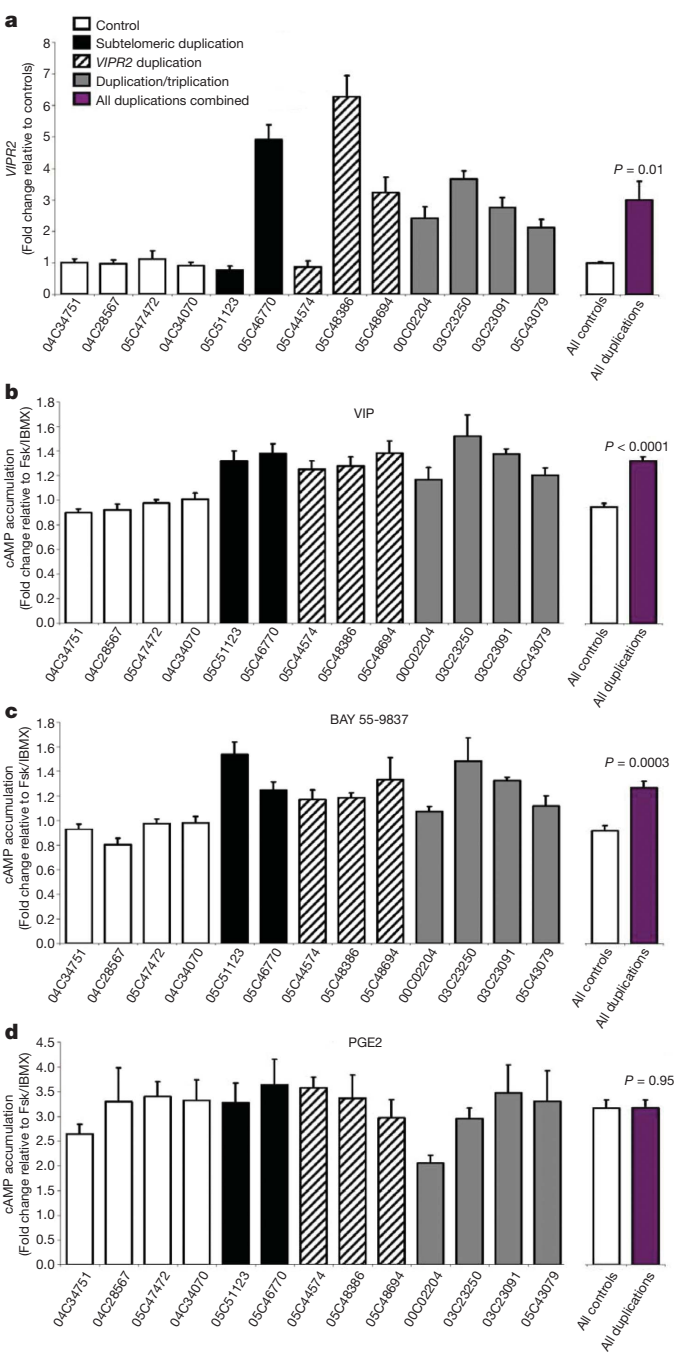
doi:10.1038/nature10088

Duplications of the neuropeptide receptor gene *VIPR2* confer significant risk for schizophrenia

Vladimir Vacic, Shane McCarthy, Dheeraj Malhotra, Fiona Murray, Hsun-Hua Chou, Aine Peoples, Vladimir Makarov, Seungtae Yoon, Abhishek Bhandari, Roser Corominas, Lilia M. Iakoucheva, Olga Krastoshevsky, Verena Krause, Verónica Larach-Walters, David K. Welsh, David Craig, John R. Kelsoe, Elliot S. Gershon, Suzanne M. Leal, Marie Dell Aquila, Derek W. Morris, Michael Gill, Aiden Corvin, Paul A. Insel, Jon McClellan, Mary-Claire King, Maria Karayiorgou, Deborah L. Levy, Lynn E. DeLisi & Jonathan Sebat

Nature 471, 499–501 (2011)

In Figure 3 of this Letter, there were labelling errors on the *x*-axis in which the 05C48386 and 05C43079 cell-line labels were both duplicated. The labels should have corresponded to 05C44574, 05C48386, 03C23091 and 05C43079. The correct figure is shown here, and this has been corrected in the HTML and PDF version of the manuscript.



CAREERS

COMMUNICATION Meet the six new columnists for Careers p.117

CAREERS BLOG The latest discussions and news on research jobs go.nature.com/z8g4a7

NATUREJOBS For the latest career listings and advice www.naturejobs.com



G. BRIAN KARAS/GETTY IMAGES

CAREER PATHS

Back to the bench

Grants and fellowship programmes can lessen the shock of re-entry for researchers who have taken a career break.

BY KENDALL POWELL

When Jane Skok stopped working in the lab in 1985, she was still using “those squeezey things to pipette”, she recalls, and the polymerase chain reaction, a DNA amplification technique, did not exist. After earning a PhD in genetics at the Imperial Cancer Research Fund in London, now called Cancer Research UK, Skok spent ten years at home taking care of her four children. She stopped reading scientific papers, and her confidence in her ability to do research fell to “zero, bottom of the class”.

But Skok desperately missed science and needed its intellectual stimulation, so to

get reacquainted with research, she undertook a master's degree in immunology at Imperial College London. During those studies, her mentor suggested that she apply for a Wellcome Trust Career Re-entry Fellowship — a programme akin to a postdoctoral fellowship but designed specifically for individuals returning to research after a significant career break. Fellows work in the laboratory of a sponsoring mentor for two to four years, which gives them time to retrain themselves, publish papers and realign themselves to apply for independent positions.

Grants and fellowship programmes aimed at encouraging talented and highly trained researchers to return to the scientific workforce

have proliferated in the past decade (see ‘A helping hand’). The vast majority of people taking advantage of these schemes are women who have had children and left the scientific pipeline after graduate degrees or postdoctoral positions. But programmes for returnees are picking up researchers of all sorts — men and women who have stepped away from the bench for reasons ranging from accommodating the career needs of spouses to trying out an alternative career or caring for parents.

Getting back to research after a long break is tough — returning researchers face younger competition, lack of confidence and new tools and technologies — but it is not impossible. Skok is now an immunologist at the New York University School of Medicine. She and others who have been successful in making a comeback say that finding the right fellowships and programmes is key. But that is not all it takes; it is also important to choose re-entry projects, mentors and fields wisely.

GETTING BACK ON TRACK

Scientists who want to get back to the bench after a break must first find ways to get back into the lab environment, as Skok did with her master's. Other strategies include volunteering in a lab for a defined period of time or taking a part-time research position.

The first step, says Carol Fishman Cohen, co-founder of the career re-entry resources company iRelaunch (www.irelaunch.com), based in Boston, Massachusetts, is to walk out of the door. “Get out of the house,” she says. “Go to conferences and lectures to meet other people in your field.”

But that does not mean accepting just any position. Ruth Ross, now a molecular pharmacologist at the University of Aberdeen, UK, sought a position as a laboratory technician after a four-year childcare break. She was turned away as over-qualified, with a PhD and two postdocs on her CV. But that rejection kept her from going backwards. The hiring investigator did her a favour, she says, by not giving her the technician job. “It's impossible to move from that into being an academic,” says Ross.

Instead, the investigator pointed her to the Wellcome Trust re-entry programme. Before applying for it, Ross spent a year teaching lab courses, reading the literature and looking for a suitable research sponsor and proposal on which to base her application.

The key, she says, was finding a project that was doable for someone coming back from a break, but competitive enough to launch a ►

► career. “I had been at home with my children. The project had to be something that would give me retraining but would also be at the forefront of research,” says Ross.

The retraining should involve minimal updating, so the work should build on the returning researcher’s existing technical skills, advises Ross. She also suggests learning about re-entry fellowships and identifying potential sponsors before taking a break, and then staying in touch with them. Ross and others say that it is vital to find a mentor who is willing to take a chance and who is supportive and flexible about the need to work part-time when starting back. Potential sponsors can be identified at conferences, or by contacting ex-colleagues or perusing job advertisements.

Knowing about a fellowship programme before taking a break can be a career saver. Allison Arnold did a postdoc at Stanford University in Palo Alto, California, where she developed a computer model to test treatments for walking disabilities. But after the postdoc, and before she was ready to enter the job market herself, Arnold moved across the country to accommodate her husband’s scientific career. A newborn kept her from the lab for two years.

“Fortunately, I had heard of the National Institutes of Health [NIH] re-entry supplements before I left Stanford,” says Arnold, who

has taken advantage of such a grant to get a job as a research associate at Harvard University’s Concord Field Station in Bedford, Massachusetts. “Without it, I would have quit my research career,” she says. Arnold is studying muscle forces in goats and guinea fowl, a project that could end up validating her post-doctoral computer model.



“Get out of the house. Go to conferences and lectures to meet others in your field.”

Carol Fishman Cohen

benefits, and includes a travel and research-expense budget. Peter Ogunbiyi, programme director in the diversity training branch of the NIH’s National Cancer Institute (NCI), notes that the NCI currently funds five re-entry supplements, and generally accepts all

applications, as long as they meet the eligibility criteria. Historically, the numbers of applications have been low.

Aspiring returnees who are not aware of re-entry programmes should stay alert for valuable opportunities. Margaret Rayman, who now researches nutritional medicine at the University of Surrey in Guildford, UK, learned of a scheme purely by chance. After a 17-year career break, Rayman had given up on getting back to science. But she heard about the Daphne Jackson Fellowship on the radio. The UK scheme, targeted at anyone returning to research from a break for family reasons, boasts a 96% success rate for fellows resuming a scientific career. Rayman used the fellowship to retrain, encouraged by a friend who worked at the University of Surrey. “The fellowship got me my credibility back — without it, I wouldn’t be a full professor,” says Rayman.

One of the biggest hurdles that Rayman faced was a lack of self-confidence — a common problem. Cohen advises returnees to meet with past colleagues, who will probably still think of them in a professional context, but warns not to pump them for recommendations or jobs; instead, she suggests asking them how research in their field has changed in the intervening years, and using the opportunity to build self-esteem. Returnees

USEFUL PROGRAMMES

A helping hand

Wellcome Trust Career Re-entry Fellowship

What: A full- or part-time fellowship of 2–4 years at an institution in Britain or Ireland. Applicants need a sponsor to provide laboratory space. Who: Researchers with at least two years of postdoctoral experience who have taken a break of two or more years and are residents of the European Economic Area or have done a degree in Britain or Ireland. How much: Provides the applicant’s salary (based on institutional standards) and research and travel expenses. Current awards are £250,000–300,000 (US\$410,000–490,000). Success rates for applications are 20–30%. go.nature.com/mjissg

Daphne Jackson Fellowship

What: A two-year, part-time fellowship to help UK researchers returning to science, technology, engineering and maths (STEM)

fields. Includes 100 hours of tailored retraining per year. Who: UK residents or people with indefinite leave to remain, who have a PhD or at least a first degree in STEM and three years of work experience, and who have taken a career break of two years or more. How much: A stipend to cover a part-time salary and £1,000 a year for research expenses not covered by the laboratory. www.daphnejackson.org

NIH Research Supplements to Promote Re-Entry into Biomedical and Behavioral Research Careers

What: Supplements to active US National Institutes of Health research grants, to support hiring people with high potential to re-enter research. The parent grant must have at least two years of funding left at the time of application. Research may be full- or part-time. Who: US citizens or permanent

residents with a doctoral-level degree who were in a postdoctoral or faculty position at the time they left research, and have had a break of 1–8 years. How much: Salary and benefits (based on institutional standards) for 1–3 years and up to \$10,000 for supplies, travel and publication costs. go.nature.com/rc3arp

Helmholtz Association re-entry positions and equal-opportunity initiatives

What: Programmes to encourage re-entry and promotion within the research workforce at the 17 Helmholtz Centres in Germany, especially for women. Each centre has its own positions. For example, in 2010, the Karlsruhe Institute of Technology had eight re-entry positions that included full salary and benefits. All full-time postdocs and faculty-level positions at Helmholtz Centres get government-paid parental

leave of up to 12 months, at 65% of normal income. go.nature.com/vjstwj

Massachusetts Institute of Technology (MIT) Career Re-engineering Program

What: A 12-month, part-time programme to help relaunch careers. Includes a one-semester MIT course, career-development workshops, an internship or research project and a final paper and presentation. Who: Applicants with at least a bachelor’s degree. How much: For 2010–11, tuition and fees cost about \$13,000. go.nature.com/fr2epe

Open University Return to Science, Engineering and Technology (SET) Course

What: A ten-week online course to assist a return to SET careers. Who: Students in Europe. How much: From £205 (in Britain) to £440 (other areas). go.nature.com/qxj1ws

should discuss their career aspirations with family and friends to rehearse for more formal interviews.

BENEFITS TO TAKING A BREAK

Spending time away from research brings many challenges, but also some advantages. Returnees, says Cohen, offer more experience and maturity than early-career scientists straight out of university, and have a renewed enthusiasm and excitement for their field. "These are huge benefits in the employer's eyes," says Cohen.

After a bachelor's degree in biology, Michelle Kwak tried her hand at several unrelated careers, but eventually got back to science through a part-time programme at the Massachusetts Institute of Technology (MIT) in Cambridge that combined MIT courses with career-development workshops and culminated in an internship.

Kwak, now a neuroscience technician at MIT, plans to apply to PhD programmes next. The straight track from bachelor's to PhD "would have been more efficient", she says, "but exploring these other areas helped me develop as a person".

David Karlin, a virologist at the University of Oxford, UK, also says that his time away improved his outlook. Karlin left research to direct a programme teaching molecular and cellular biology to the public at a university in France. During his seven-year break, he kept up research with collaborators, which helped with his Wellcome fellowship application.

Karlin's outreach work gave him a broader perspective and vision for his research. "I was hired at Oxford precisely because I had an original approach," he says. "I cannot hide my atypical career path, so, as a colleague says, 'When you have a wooden leg, shake it!'"

Skok and others encourage young researchers to consider a break, especially if they become stretched between demanding research and responsibilities at home. Despite the challenges associated with returning to research after a long hiatus, many researchers have made the transition with the right planning, and the rewards are huge — especially for those who are passionate about the scientific process.

After several career stops and starts — necessary to take care of a son who needed two heart transplants — Ana Hoffenberg, a medical doctor, gravitated towards research. Driven by her intellectual curiosity and passion for knowledge, she eventually landed a postdoc fellowship at the University of Colorado in Denver. "I'm hoping this will be the opening door to an independent research career," she says. "With the hospital bills, I needed income. But I also needed the playground in my mind." ■

Kendall Powell is a freelance science writer based in Lafayette, Colorado.

COMMUNICATION

Voices for Careers

Columnist competition yields six winners.



BY KAREN KAPLAN

Nature would like to introduce our six new Careers columnists, the winners of an international contest.

Earlier this year, we sought essayists immersed in the challenges of graduate studies and postdoctoral fellowships — stages that are crucial to fledgling researchers. We selected three doctoral students and three postdocs from four continents, who conduct research in the diverse fields of synthetic chemistry, genetics, civil and environmental engineering, ecology, pain research and cell biology. Half of them work in the United States, and half in other countries. Choosing the winners was even more daunting than for our past contests; we received nearly 300 entries from close to 40 countries.

Competition was fierce and we chose an elite group of writers who have not only outstanding communication skills, but also a variety of backgrounds and interests. The winning entries were incisive, contemplative and clever — and told compelling stories.

The columns, and shorter journal entries, will run online at go.nature.com/uy5nyc, with a few columns appearing periodically in the section's print edition.

A PROBLEM SHARED

Newly minted Careers columnist Adam James, a PhD candidate in synthetic chemistry at the University of Tasmania in Hobart, Australia, is motivated to write because he enjoys outlining problems and offering answers. James, who took philosophy courses as an undergraduate, is proud of his ability to construct an argument, and hopes to offer meaningful and valid solutions to common concerns encountered by our readers.

Lucie Low, originally from the United

Kingdom, is a postdoc studying the neuroscience of chronic pain at McGill University in Montreal, Canada. In her sample column entry, she brought both a keen sense of humour and a serious outlook to a discussion of Canadian laws that tax postdoc stipends yet leave postdoctoral researchers with minimal employment benefits.

First-year PhD student Lydia Murray also has a wry sense of humour. In her writing, she wants to help others come to terms with the 'lows' of research. For Murray, who is studying molecular genetics and cell biology at the University of Glasgow, UK, the importance of buoying others' spirits is one of the most important lessons from her first year of graduate study. Giving a forlorn colleague a pat on the back might assist them more than is immediately obvious, she writes.

COMMUNICATION OF ALL SORTS

Mariano Loza-Coll, now in his second postdoc position, sees communication as science's biggest single challenge. A native of Argentina, Loza-Coll works at the Salk Institute for Biological Studies in La Jolla, California, and would like to promote more effective transmission of ideas between scientists, the media and the public. He hopes to meet this goal through his research and writing.

Andrew Peterman offers the perspective of an academic who has ties to the corporate world. A PhD candidate in civil and environmental engineering at Stanford University in California, Peterman has researched and devised strategies to help Walt Disney Imagineering in Glendale, California, to reduce energy consumption. That taught him a lesson about communication: it is often a challenge, even among colleagues.

Communicating science, in this case to students, is also an interest of Gaston Small, a postdoc in ecology at the University of Minnesota in St Paul. Small, who has a PhD in ecology and a master's degree in teaching with a focus on science education, is investigating the role of microbes in increased nitrate concentrations in Lake Superior. With two young children, Small is also addressing the challenges of a work-life balance from a father's point of view. He is in the process of deciding whether to stay in research or return to the classroom.

We hope that readers will track our columnists' progress, points of view and perspectives with interest as they pursue their own career aims. And we offer our sincere thanks to all who applied. ■

L. ALLEN/CORBIS

should discuss their career aspirations with family and friends to rehearse for more formal interviews.

BENEFITS TO TAKING A BREAK

Spending time away from research brings many challenges, but also some advantages. Returnees, says Cohen, offer more experience and maturity than early-career scientists straight out of university, and have a renewed enthusiasm and excitement for their field. "These are huge benefits in the employer's eyes," says Cohen.

After a bachelor's degree in biology, Michelle Kwak tried her hand at several unrelated careers, but eventually got back to science through a part-time programme at the Massachusetts Institute of Technology (MIT) in Cambridge that combined MIT courses with career-development workshops and culminated in an internship.

Kwak, now a neuroscience technician at MIT, plans to apply to PhD programmes next. The straight track from bachelor's to PhD "would have been more efficient", she says, "but exploring these other areas helped me develop as a person".

David Karlin, a virologist at the University of Oxford, UK, also says that his time away improved his outlook. Karlin left research to direct a programme teaching molecular and cellular biology to the public at a university in France. During his seven-year break, he kept up research with collaborators, which helped with his Wellcome fellowship application.

Karlin's outreach work gave him a broader perspective and vision for his research. "I was hired at Oxford precisely because I had an original approach," he says. "I cannot hide my atypical career path, so, as a colleague says, 'When you have a wooden leg, shake it!'"

Skok and others encourage young researchers to consider a break, especially if they become stretched between demanding research and responsibilities at home. Despite the challenges associated with returning to research after a long hiatus, many researchers have made the transition with the right planning, and the rewards are huge — especially for those who are passionate about the scientific process.

After several career stops and starts — necessary to take care of a son who needed two heart transplants — Ana Hoffenberg, a medical doctor, gravitated towards research. Driven by her intellectual curiosity and passion for knowledge, she eventually landed a postdoc fellowship at the University of Colorado in Denver. "I'm hoping this will be the opening door to an independent research career," she says. "With the hospital bills, I needed income. But I also needed the playground in my mind." ■

Kendall Powell is a freelance science writer based in Lafayette, Colorado.

COMMUNICATION

Voices for Careers

Columnist competition yields six winners.



BY KAREN KAPLAN

Nature would like to introduce our six new Careers columnists, the winners of an international contest.

Earlier this year, we sought essayists immersed in the challenges of graduate studies and postdoctoral fellowships — stages that are crucial to fledgling researchers. We selected three doctoral students and three postdocs from four continents, who conduct research in the diverse fields of synthetic chemistry, genetics, civil and environmental engineering, ecology, pain research and cell biology. Half of them work in the United States, and half in other countries. Choosing the winners was even more daunting than for our past contests; we received nearly 300 entries from close to 40 countries.

Competition was fierce and we chose an elite group of writers who have not only outstanding communication skills, but also a variety of backgrounds and interests. The winning entries were incisive, contemplative and clever — and told compelling stories.

The columns, and shorter journal entries, will run online at go.nature.com/uy5nyc, with a few columns appearing periodically in the section's print edition.

A PROBLEM SHARED

Newly minted Careers columnist Adam James, a PhD candidate in synthetic chemistry at the University of Tasmania in Hobart, Australia, is motivated to write because he enjoys outlining problems and offering answers. James, who took philosophy courses as an undergraduate, is proud of his ability to construct an argument, and hopes to offer meaningful and valid solutions to common concerns encountered by our readers.

Lucie Low, originally from the United

Kingdom, is a postdoc studying the neuroscience of chronic pain at McGill University in Montreal, Canada. In her sample column entry, she brought both a keen sense of humour and a serious outlook to a discussion of Canadian laws that tax postdoc stipends yet leave postdoctoral researchers with minimal employment benefits.

First-year PhD student Lydia Murray also has a wry sense of humour. In her writing, she wants to help others come to terms with the 'lows' of research. For Murray, who is studying molecular genetics and cell biology at the University of Glasgow, UK, the importance of buoying others' spirits is one of the most important lessons from her first year of graduate study. Giving a forlorn colleague a pat on the back might assist them more than is immediately obvious, she writes.

COMMUNICATION OF ALL SORTS

Mariano Loza-Coll, now in his second postdoc position, sees communication as science's biggest single challenge. A native of Argentina, Loza-Coll works at the Salk Institute for Biological Studies in La Jolla, California, and would like to promote more effective transmission of ideas between scientists, the media and the public. He hopes to meet this goal through his research and writing.

Andrew Peterman offers the perspective of an academic who has ties to the corporate world. A PhD candidate in civil and environmental engineering at Stanford University in California, Peterman has researched and devised strategies to help Walt Disney Imagineering in Glendale, California, to reduce energy consumption. That taught him a lesson about communication: it is often a challenge, even among colleagues.

Communicating science, in this case to students, is also an interest of Gaston Small, a postdoc in ecology at the University of Minnesota in St Paul. Small, who has a PhD in ecology and a master's degree in teaching with a focus on science education, is investigating the role of microbes in increased nitrate concentrations in Lake Superior. With two young children, Small is also addressing the challenges of a work-life balance from a father's point of view. He is in the process of deciding whether to stay in research or return to the classroom.

We hope that readers will track our columnists' progress, points of view and perspectives with interest as they pursue their own career aims. And we offer our sincere thanks to all who applied. ■

L. ALLEN/CORBIS

PRIVATE EXPLORATION

A crude result.

BY T. C. MCCARTHY

*I*t had to be crude oil, Terrell thought. He adjusted the sample again and then his face mask, forcing himself to stay calm. The shimmer was there, the rock like one of the shales he had studied in Sweden, except that Terrell knew it came from someplace much farther away, and the shimmer was what everyone wanted to find, but which nobody dreamed they would. Terrell switched to ultraviolet light and the rock glowed in confirmation as if smiling back, daring him to throw all the tests at it he could, that it wasn't scared and he should just go ahead. "I double dare you," it would have said — if rocks could talk.

Oil! On Mars!

Margaret sat in the next station, her hair covered and a mask hiding everything except her nose and glasses. Terrell watched for a minute, not sure what to say.

"Margaret, where did my rock come from again? T-one-seven-one-alpha?"

"Mars," she said dryly.

"Ha, ha. Seriously. Which mission?" There had been three separate core-return missions that just got back, and Terrell didn't really have to get an answer from Margaret, he could have looked for himself. But it was the first thing that popped into his head. Margaret asked for the number again and then checked.

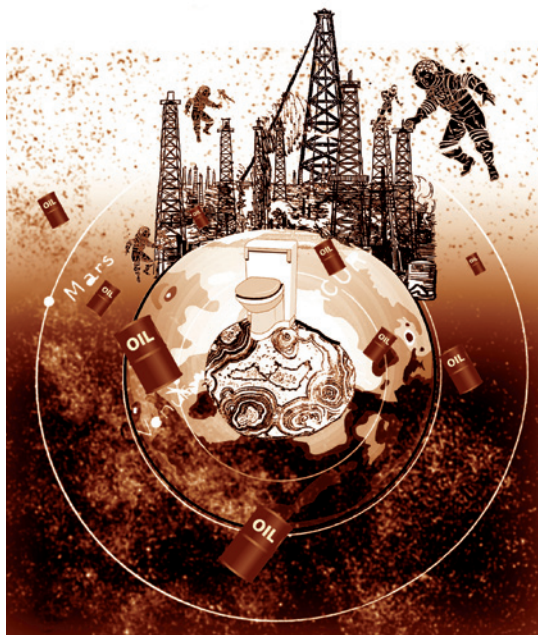
"Mission Two, Midland Petroleum, and your sample's from the core bottom, the deepest section. Everything else from that one was evaporite. Evaporite, evaporite, evaporite, just like Mission One cores, just like Mission Three's, and just like every single mission before if you don't count all the basalt. Why? Find any oil?"

Terrell almost said yes, but it was the same sarcasm she always used and he laughed it off, instead saying something like, *yeah, a whole barrel of it*. "Ever think about what would happen?"

"What would happen where? When?"

Terrell turned off his microscope. "If we found oil. On Mars."

"Yeah, now that you mention it." She became animated, and he saw an angry glare in her eyes, one that mirrored her voice. "I think about it all the time. The day Mars was finally colonized, but not by explorers. By Midland and a hundred other corporations who killed each other to buy up all the



land rights, and then sent company geologists to drill, company executives hopping a ride just so they could say that they'd been to Mars. And the drillers. Don't get me started on drillers, I don't want to know what they'd do without their brothels. Think about it. When would anyone let colonists onto Mars in that scenario?"

Terrell thought but couldn't answer, and then shook his head. "When?"

"Just as soon as Mars was destroyed, covered with muck piles, empty mud pits, and spilled oil. After nobody *wants* to see Mars."

"It wouldn't be that bad," said Terrell. But already he wished he hadn't said anything.

"Really? I'm from west Texas. Ever been to west Texas?"

"No. Never been west of Houston."

Margaret pointed at him as if she had triumphed. "Exactly."

"So why are you working here?" he asked. "Why bother?"

"Same reason as you." Margaret returned to her microscope, adjusting the light and magnification before she looked back a moment later. "Where else can a petroleum geologist find work these days?"

"Yeah," said Terrell. "Where else?"

The smile had abandoned him. In fact, his excitement had also drained completely, leaving only a sickening sensation that once again he hadn't thought

things through, hadn't considered the ramifications of what he was doing, because what if Margaret was right? Terrell's shoulder patch said Corchoran. It was short for Corchoran Diversified, a services company that contracted for just about everything, and once the professorships hadn't panned out, it was the only job Terrell had found, the only one that offered insurance, the only geology job anywhere — subcontracting sample-return analyses for the Mars missions, all of them privately funded. But these would be the last three missions and once the analyses were done, he'd hit the street again for interview after interview until he finally wound up delivering pizzas. The truth of it hit him then: Corchoran didn't give a damn either.

There was no telling if the cameras were watching. They were always on, always pointed at them from inside their sterile booths, and as soon as he decided what he'd do, Terrell felt sweat on his forehead.

It was a 50–50 shot; the guard would either be watching him or Margaret. He opened the sample container and pretended to put the rock inside, palming it instead, and once outside the booth, with Margaret calling out to ask where he was going, he stepped into the prep room and began taking off his gloves, booties and apron. Terrell slipped the rock into a shoe, deep, where he felt it under his foot. He did his best not to limp. Less than a minute later he had got through the pocket search, the guard paying only half attention to his explanation that he needed to use the can, before Terrell slipped into the bathroom. He shut the stall door and pulled his shoes off, and stared at the rock one last time. It only took a second to drop it in the toilet. Terrell flushed twice. He didn't bother to get his things then, but instead left through the front door in socks. It was all Terrell could do to stop thinking about what he had just thrown away, especially once he realized they'd probably arrest him. But those thoughts didn't last long. Four words took their place, running through his mind like a steady drum beat, over and over until he grinned.

Crude oil. On Mars! ■

T. C. McCarthy's fiction has appeared in *Per Contra* and will soon appear in *Story Quarterly*. His debut novel, *Germline*, is forthcoming in 2011 from Orbit Books. Visit T. C. at www.tcmccarthy.com.

➤ **NATURE.COM**
Follow Futures on
Facebook at:
go.nature.com/mtoodm

Mineralogical constraints on Precambrian p_{CO_2} ARISING FROM M. T. Rosing, D. K. Bird, N. H. Sleep & C. J. Bjerrum *Nature* **464**, 744–747 (2010)

Rosing and others¹ recently proposed a new model for the early Earth's atmospheric composition. They suggest, on the basis of mineral assemblages in sedimentary rocks, that atmospheric CO_2 and CH_4 concentrations have been consistently overestimated. This proposal is intriguing, given that high concentrations of these gases are traditionally considered to be the solution to the 'faint young Sun paradox' of why the Earth did not freeze over in our Sun's youth when solar luminosity was greatly decreased. Rosing and others¹ instead invoke a decreased planetary albedo during the Earth's early history. New insights into this problem, which has long captivated those interested in the Earth's early climate and biological history, are welcome. However, we believe that the arguments presented by Rosing and others¹ for roughly modern values of the partial pressure of CO_2 (p_{CO_2}) are not robust.

A fundamental assumption in the model of Rosing and others¹ is that the coexistence of the mineral phases siderite (FeCO_3) and magnetite (Fe_3O_4) in banded iron formations (BIFs) represents an assemblage that is nearly in thermodynamic equilibrium with the atmosphere, placing stringent constraints on the partial pressures of both CO_2 and H_2 . The mineralogy of some BIFs is dominated by FeCO_3 (ref. 2), whereas others have mineralogies dominated by ferric oxide phases³. However, in well-studied BIFs with mixed Fe_3O_4 – FeCO_3 phases (such as the Kuruman, Hamersley and Old Wanderer iron formations), carbon and iron isotope work^{4,5} indicates that these minerals did not precipitate in isotopic equilibrium with the ocean. Therefore, diagenetic conditions controlled mineral formation; there is evidence for an initial rain of ferric oxides and secondary formation of reduced and mixed-valence iron minerals during early or later-stage diagenesis^{4,5}. This is important given that dissimilatory iron-reducing bacteria generate Fe_3O_4 even at very high aqueous carbon dioxide concentrations and headspace values of $[\sum \text{CO}_2] \approx 50 \text{ mM}$ (ref. 6) and $p_{\text{CO}_2} \approx 0.2 \text{ atm}$ (ref. 7). Mixed Fe_3O_4 – FeCO_3 assemblages develop as a result of an imbalance between rates of non-reductive ferric oxide dissolution, Fe^{2+} transport, and rates of Fe^{3+} – Fe^{2+} conversion^{6–8}. Therefore, p_{CO_2} values very much above the modern value do not preclude the formation of magnetite.

In addition, the authors assume that both the atmospheric partial pressure and aqueous concentration of H_2 will be controlled by hydrogenotrophic methanogenesis. However, the preservation of ferric oxides, together with a paucity of organic carbon in most BIFs, strongly suggests that H_2 pressures or concentrations would instead be buffered most often by dissimilatory iron-reducing bacteria in BIF diagenetic environments. This would decrease the aqueous concentration of H_2 significantly⁹, changing the relevant stability boundary to Fe_2O_3 – FeCO_3 . For instance, a threshold $[\text{H}_2]$ of about 0.1 nM would yield p_{CO_2} estimates between about 30–100 PAL (present atmospheric level) at 25 °C–35 °C (Fig. 1a of ref. 1). We view this as a non-trivial difference given the attendant implications for climatologically plausible CH_4/CO_2 ratios and atmospheric CH_4 concentrations¹⁰.

Finally, even if we were to assume a simple thermodynamic control regulated by p_{H_2} (via methanogenesis) and p_{CO_2} , it is unlikely that the coexistence of Fe_3O_4 and FeCO_3 in BIFs provides any direct evidence regarding the chemistry of a surface ocean that is nearly in equilibrium with the overlying atmosphere. There is overwhelming petrographic evidence that the Fe_3O_4 observed in many BIFs is metamorphic or late-stage alteration product^{11–13}. Indeed, the coexistence of Fe_2O_3 – Fe_3O_4 – FeCO_3 assemblages and significant mineralogical variability on small spatial scales in many BIFs suggest

either the formation and preservation of mineral phases out of thermodynamic equilibrium with the ambient ocean–atmosphere system or secondary/metamorphic alteration. Although we do not suggest that a single depositional model applies for all BIF occurrences throughout the Earth's history, we do contend that many of the basic processes operating during their formation (microbial Fe oxidation/reduction, organic-matter remineralization in sediments, and metamorphism) are pervasive if not ubiquitous and should be considered in any attempt to relate BIF mineralogy to atmospheric composition.

Rosing *et al.*¹ rightly point out that some previous work (see references in ref. 1) based on mineral stability in palaeosol profiles has also suggested low p_{CO_2} values. However, this work contrasts with experimental studies¹⁴, indicating that mineral stability as a function of p_{CO_2} during the formation of ancient soil profiles is not fully understood. Nonetheless, the numerical model proposed by Rosing *et al.*¹ has a number of important implications. For example, by implicitly linking the Earth's planetary albedo to the oxidation state of the atmosphere, the model may provide a new mechanism for explaining the inception of widespread glaciation following changes in Proterozoic atmospheric oxygen content. Such considerations are novel and will be fruitful to examine, and it is likely that any explanation for the clement early Earth will involve a complex interplay of forcings and feedbacks. However, we conclude that early Precambrian atmospheric p_{CO_2} levels remain poorly constrained and that increased levels of atmospheric gases (CO_2 , CH_4 , C_2H_2) remain a compelling solution to the 'faint young Sun paradox'.

Christopher T. Reinhard¹ & Noah J. Planavsky¹

¹Department of Earth Sciences, University of California, Riverside, Riverside, California, USA.
e-mail: crein003@ucr.edu

Received 13 April 2010; accepted 26 January 2011.

- Rosing, M. T., Bird, D. K., Sleep, N. H. & Bjerrum, C. J. No climate paradox under the faint early Sun. *Nature* **464**, 744–747 (2010).
- Goodwin, A. M. Geochemical studies at the Helen iron range. *Econ. Geol.* **59**, 684–718 (1964).
- Hofmann, A. The geochemistry of sedimentary rocks from the Fig Tree Group, Barberton greenstone belt: implications for tectonic, hydrothermal and surface processes during mid-Archaean times. *Precamb. Res.* **143**, 23–49 (2005).
- Johnson, C. M., Beard, B. L., Klein, C., Beukes, N. J. & Roden, E. E. Iron isotopes constrain biologic and abiologic processes in banded iron formation genesis. *Geochim. Cosmochim. Acta* **72**, 151–169 (2008).
- Steinheofel, G., Horn, I. & von Blanckenburg, F. Micro-scale tracing of Fe and Si isotope signatures in banded iron formation using femtosecond laser ablation. *Geochim. Cosmochim. Acta* **73**, 5343–5360 (2009).
- Behrends, T. & Van Cappellen, P. Transformation of hematite into magnetite during dissimilatory iron reduction—conditions and mechanisms. *Geomicrobiol. J.* **24**, 403–416 (2007).
- Roh, Y. *et al.* Biogeochemical and environmental factors in Fe biomineralization: magnetite and siderite formation. *Clays Clay Miner.* **51**, 83–95 (2003).
- Zachara, J. M., Kukkadapu, R. K., Fredrickson, J. K., Gorby, Y. A. & Smith, S. C. Biomineralization of poorly crystalline Fe(III) oxides by dissimilatory metal reducing bacteria (DMRB). *Geomicrobiol. J.* **19**, 179–207 (2002).
- Lovley, D. R. & Goodwin, S. Hydrogen concentrations as an indicator of the predominant terminal electron-accepting reactions in aquatic sediments. *Geochim. Cosmochim. Acta* **52**, 2993–3003 (1988).
- Haqq-Misra, J. D., Domagal-Goldman, S. D., Kasting, P. J. & Kasting, J. F. A revised, hazy methane greenhouse for the Archean Earth. *Astrobiology* **8**, 1127–1137 (2008).
- Han, T.-M. Origin of magnetite in Precambrian iron-formations of low metamorphic grade. *Proc. 7th Quadrennial IAGOD Symp.* (ed. Zachrisson, E.) 641–656 (1988).

12. Krapež, B., Barley, M. E. & Pickard, A. L. Hydrothermal and resedimented origins of the precursor sediments to banded iron formations: sedimentological evidence from the early Palaeoproterozoic Brockman Supersequence of Western Australia. *Sedimentology* **50**, 979–1011 (2003).
13. Tompkins, L. A. & Cowan, D. R. Opaque mineralogy and magnetic properties of selected banded iron-formations, Hamersley Basin, Western Australia. *Aust. J. Earth Sci.* **48**, 427–437 (2001).

14. Murakami, T. *et al.* Anoxic dissolution processes of biotite: implications for Fe behavior during Archean weathering. *Earth Planet. Sci. Lett.* **224**, 117–129 (2004).

Author Contributions C.T.R. and N.J.P. both wrote the article.

Competing financial interests: declared none.

doi:10.1038/nature09959

Low p_{CO_2} in the pore water, not in the Archean air

ARISING FROM M. T. Rosing, D. K. Bird, N. H. Sleep & C. J. Bjerrum *Nature* **464**, 744–747 (2010)

The solar luminosity during the Archean (3.8 to 2.5 billion years ago) was 20–25% lower than at present, and was probably compensated, at least in part, by a stronger greenhouse effect^{1–3}. Rosing *et al.*⁴ estimate the Archean partial pressure of carbon dioxide to have been about $p_{\text{CO}_2} \approx 10^{-3}$ bar, on the basis of the simultaneous occurrence of magnetite (Fe_3O_4) and siderite (FeCO_3) in banded iron formations (BIFs, a type of chemical sediment). Here, we question a central assumption by Rosing *et al.*⁴ that the mineralogy of BIFs reflects near-thermodynamic equilibrium with the atmosphere–ocean system; just as the presence of authigenic pyrite in modern sediments does not imply that the modern atmosphere is anoxic, the mineralogy of BIFs cannot be used to argue for a low- p_{CO_2} Archean atmosphere. Carbon dioxide is therefore still a viable greenhouse gas candidate with which to explain the warm Archean climate.

The original iron precipitate in the BIFs was probably ferrihydrite, $\text{Fe}_2\text{O}_3 \cdot x\text{H}_2\text{O}$. The ferrihydrite was precipitated out of equilibrium with the atmosphere by oxidation of upwelled ferrous iron, either by O_2 produced by cyanobacteria within the water column or by phototrophic, anoxygenic, iron-oxidizing bacteria. Once in the sediment, ferrihydrite was subsequently converted to a stable iron oxide, either magnetite or haematite, or to the reduced mineral siderite. The constraint on p_{H_2} – p_{CO_2} derives from applying the following equilibrium reaction between magnetite and siderite: $\text{Fe}_3\text{O}_4 + \text{H}_2 + 3\text{CO}_2 \leftrightarrow 3\text{FeCO}_3 + \text{H}_2\text{O}$. As discussed below, p_{H_2} is a measure of the redox potential of the system, but organic matter (CH_2O) was the reducing agent in this reaction.

In the model of Rosing *et al.*⁴, p_{H_2} was controlled by methanogens. Anaerobic ecosystems of this nature have been studied by ref. 5. The downward flux of H_2 through the atmosphere–ocean interface is limited by its piston velocity to $(1\text{--}6) \times 10^{11}$ molecules $\text{cm}^{-2} \text{s}^{-1}$ (Table 2 in ref. 5). The average deposition rate of Fe_3O_4 in BIFs is estimated to be $0.1\text{--}1 \text{ mm yr}^{-1}$, or $(40\text{--}400) \times 10^{11}$ molecules $\text{cm}^{-2} \text{s}^{-1}$, assuming that each microband represents one year of deposition⁶. Suggestions that the deposition rate was much slower than this^{7,8} are probably biased by hiatuses in the geologic record. Reducing this iron to siderite would require an equal flux of H_2 , which is 7–400 times the downward H_2 flux into the ocean estimated above. Some H_2 could have been produced within the sediment via fermentation, but this requires that organic matter be available. Formation of siderite without a reducing agent—that is, by disproportionation ($\text{Fe}_3\text{O}_4 + \text{CO}_2 \rightarrow \text{Fe}_2\text{O}_3 + \text{FeCO}_3$)—would require an effective p_{H_2} of 3×10^{-6} (the value at the magnetite–haematite boundary), which is ten times smaller than the value that Rosing *et al.*⁴ propose and is probably impossible to achieve.

Probably the reducing agent was not H_2 but organic matter produced by photosynthesis in the water column. A fraction of the primary production of organic matter was exported from the photic zone to the sediment, where microbial iron respiration could produce siderite^{9–12}: $2\text{Fe}_3\text{O}_4 + \text{CH}_2\text{O} + 5\text{CO}_2 \rightarrow 6\text{FeCO}_3 + \text{H}_2\text{O}$. Light carbon isotope ratios

are found in siderite, but not in associated Ca/Mg carbonates^{11–13}, indicating that a portion of the carbon in siderite was indeed derived from a pool of isotopically light organic matter. In BIFs, magnetite was initially in disequilibrium with the overlying CO_2 -rich atmosphere and was converted to siderite. However, the above reaction was only allowed to proceed as long as a reductant (organic matter) was available. The rain of organic carbon into the sediment was not sufficient to keep pace with that of Fe^{3+} -bearing oxide, and its exhaustion led to a mineral assemblage (coexisting magnetite and siderite) that did not reflect equilibrium conditions with the atmosphere–ocean system. The transition from siderite to iron oxides in BIFs from the Campbellrand–Kuruman complex could have been caused by the decreased supply of organic material in offshore regions¹⁰.

To summarize, microbial cycling and diagenesis of organic matter and ferric iron controlled the effective p_{H_2} and p_{CO_2} in the sediment. The mineralogy of BIFs reflects these conditions, with no simple relationship to atmospheric CO_2 . Thus, p_{CO_2} in the atmosphere may have been high enough (about 0.1 bar) to explain the warm Archean climate, without the need for additional warming mechanisms.

N. Dauphas¹ & J. F. Kasting²

¹Origins Laboratory, Department of the Geophysical Sciences and Enrico Fermi Institute, The University of Chicago, Chicago, Illinois 60637, USA. e-mail: dauphas@uchicago.edu

²Department of Geosciences, The Pennsylvania State University, University Park, Pennsylvania 16802, USA.

Received 1 June 2010; accepted 26 January 2011.

- Sagan, C. & Mullen, G. Earth and Mars: evolution of atmospheres and surface temperatures. *Science* **177**, 52–56 (1972).
- Owen, T., Cess, R. D. & Ramanathan, V. Enhanced CO_2 greenhouse to compensate for reduced solar luminosity on early Earth. *Nature* **277**, 640–642 (1979).
- Walker, J. C. G., Hays, P. B. & Kasting, J. F. A negative feedback mechanism for the long-term stabilization of Earth's surface temperature. *J. Geophys. Res.* **86**, 9776–9782 (1981).
- Rosing, M. T., Bird, D. K., Sleep, N. H. & Bjerrum, C. J. No climate paradox under the faint early Sun. *Nature* **464**, 744–747 (2010).
- Kharecha, P., Kasting, J. & Siefert, J. A coupled atmosphere–ecosystem model of the early Archean Earth. *Geobiology* **3**, 53–76 (2005).
- Konhauser, K. O. *et al.* Could bacteria have formed the Precambrian banded iron formations? *Geology* **30**, 1079–1082 (2002).
- Pickard, A. L. SHRIMP U–Pb zircon ages for the Palaeoproterozoic Kuruman iron formation, Northern Cape Province, South Africa: evidence for simultaneous BIF deposition on Kaapvaal and Pilbara cratons. *Precamb. Res.* **125**, 275–315 (2003).
- Trendall, A. F., Compston, W., Nelson, D. R., De Laeter, J. R. & Bennett, V. C. SHRIMP zircon ages constraining the depositional chronology of the Hamersley Group, Western Australia. *Aust. J. Earth Sci.* **51**, 621–644 (2004).
- Walker, J. C. G. Suboxic diagenesis in banded iron formations. *Nature* **309**, 340–342 (1984).
- Fischer, W. W. & Knoll, A. H. An iron shuttle for deepwater silica in Late Archean and early Paleoproterozoic iron formation. *Geol. Soc. Am. Bull.* **121**, 222–235 (2009).
- Heimann, A. *et al.* Fe, C, and O isotope compositions of banded iron formation carbonates demonstrate a major role for dissimilatory iron reduction in ~2.5 Ga marine environments. *Earth Planet. Sci. Lett.* **294**, 8–18 (2010).

12. Craddock, P. R. & Dauphas, N. Iron and carbon isotope evidence for microbial iron respiration throughout the Archean. *Earth Planet. Sci. Lett.* **303**, 121–132 (2011).
13. Becker, R. H. & Clayton, R. N. Carbon isotopic evidence for the origin of a banded iron-formation in Western Australia. *Geochim. Cosmochim. Acta* **36**, 577–595 (1972).

Author Contributions Both authors contributed equally to this work.

Competing financial interests: declared none.

doi:10.1038/nature09960

Faint young Sun paradox remains

ARISING FROM M. T. Rosing, D. K. Bird, N. H. Sleep & C. J. Bjerrum, *Nature* **464**, 744–747 (2010)

The Sun was fainter when the Earth was young, but the climate was generally at least as warm as today; this is known as the ‘faint young Sun paradox’. Rosing *et al.*¹ claim that the paradox can be resolved by making the early Earth’s clouds and surface less reflective. We show that, even with the strongest plausible assumptions, reducing cloud and surface albedos falls short by a factor of two of resolving the paradox. A temperate Archean climate cannot be reconciled with the low level of CO₂ suggested by Rosing *et al.*¹; a stronger greenhouse effect is needed.

During the Archean eon, the Earth received 76% to 83% of the energy from the Sun that it does today. If the Earth’s greenhouse effect and albedo were the same as now, the Earth would have been in continual deep freeze until one billion years ago, with glaciers reaching the Equator. However, Archean glacial sediments are rare and geological evidence indicates that the Archean was typically warmer than today (we are in a glacial period now). With the amount of energy reaching the Earth given by $F = \frac{1}{4}S_0(1 - \alpha) = 239 \text{ W m}^{-2}$ (using the present-day solar constant $S_0 = 1,368$ and albedo $\alpha = 0.3$), the radiative deficit in the Archean would have been $(1 - 0.79)F \approx 50 \text{ W m}^{-2}$. Resolution of the ‘faint young Sun paradox’ requires a positive radiative forcing—from reducing the albedo or increasing the greenhouse effect—of more than 50 W m^{-2} .

Clouds have two competing radiative effects: they reflect sunlight but they also add to the greenhouse effect if they are colder than the surface. Reflection dominates in low clouds, and the greenhouse effect dominates in high clouds. Therefore the absolute upper bound on warming by decreasing cloud reflectivity would be found by removing low clouds entirely. This gives a forcing of 25 W m^{-2} , half of what is needed to resolve the ‘faint young Sun paradox’ (our cloud model is described in the Methods). Any reduction to high clouds would cause a cooling.

Rosing *et al.*¹ justify less-reflective clouds with the incorrect statements that most cloud condensation nuclei (CCN) are from biogenic dimethyl sulphide (DMS), and that DMS is solely produced by eukaryotes. DMS is also produced microbially². Products of DMS contribute only 3% of Northern Hemisphere CCN and 10% of Southern Hemisphere CCN today³. Other biological⁴ and non-biological sources, especially sea salt, provide CCN. If CCN production were to depend only on eukaryotic DMS emissions¹, we would expect to see significant cooling when eukaryotes evolved, but no such cooling is evident.

Nevertheless, we can assume no biological CCN supply and quantify the resulting forcing. Over the modern ocean the effective radius r_e of cloud drops rarely exceeds $15 \mu\text{m}$ (ref. 5) even in remote and unproductive regions (the r_e of $17 \mu\text{m}$ to $30 \mu\text{m}$ used by Rosing *et al.*¹ is too high). For an upper bound, we increase low cloud droplet size by 50% from our standard case, from $11 \mu\text{m}$ to $16.5 \mu\text{m}$. With no change in cloud thickness, the forcing is 7 W m^{-2} . Clouds with larger drops may rain out faster. Parameterizations of enhanced rain-out vary from proportional to $(r_{e,0}/r_e)^1$ to proportional to $(r_{e,0}/r_e)^{5.37}$ (refs 6 and 7); the corresponding extra forcing would be $4\text{--}15 \text{ W m}^{-2}$ (remote sensing

data for marine stratus suggest that the low end of this range is more appropriate⁸). The sum is 11 W m^{-2} to 22 W m^{-2} , with the low end being most likely.

The authoritative estimate of the global energy budget⁹ gives global mean and ocean albedos of 0.125 and 0.090 respectively. The largest realistic surface darkening is from the present mean to an all-ocean world, which gives a radiative forcing of 5 W m^{-2} .

Increasing the CO₂ mixing ratio to 1,000 parts per million by volume (p.p.m.v.; the upper bound according to Rosing *et al.*¹) gives a forcing of 6 W m^{-2} . Rosing *et al.*¹ rely on 1,000 p.p.m.v. CH₄ for much of their warming, ignoring relevant atmospheric chemistry. As the partial pressure of CH₄ (p_{CH_4}) approaches that of CO₂ (p_{CO_2}), hydrocarbon haze forms in the stratosphere, the cooling effect of which outweighs the greenhouse effect of CO₂ and CH₄ (refs 10 and 11). Numerical models¹² predict haze production when $p_{\text{CH}_4}/p_{\text{CO}_2} = 0.1$ and haze production has been seen in laboratory experiments¹³ where $p_{\text{CH}_4}/p_{\text{CO}_2} = 0.3$. With 1,000 p.p.m.v. CO₂, the maximum CH₄ concentration that can give warming is 300 p.p.m.v., which would contribute 7 W m^{-2} of additional forcing.

Changes to clouds could in theory considerably reduce the amount of greenhouse gases required, because gaseous absorption depends on the logarithm of gas abundance. But even with the highly unlikely assumption of no biological CCN supply, cloud changes can provide only one-quarter to one-half of the required radiative forcing. Any changes to clouds would require strong justification, which Rosing *et al.*¹ do not provide. A strong greenhouse effect is required in the Archean. The alternative is an extremely cold climate with continual mid- to low-latitude glaciation, for which there is no evidence.

METHODS

We calculate the radiative forcing (change in net flux at the tropopause) on a single global annual mean atmospheric profile, with three layers of clouds that overlap randomly¹⁴. The radiative fluxes on eight sub-columns corresponding to each cloud combination are calculated with the RRTM model¹⁵. For our standard case, cloud water paths are $[W_{\text{high}}, W_{\text{mid}}, W_{\text{low}}] = [20, 25, 40] \text{ g m}^{-2}$, fractions are $[f_{\text{high}}, f_{\text{mid}}, f_{\text{low}}] = [0.25, 0.25, 0.40]$ and the surface albedo is 0.125. Standard low- and mid-level clouds are liquid with $r_e = 11 \mu\text{m}$ and high clouds are ice with generalized effective size of $D_{\text{ge}} = 75 \mu\text{m}$. For radiative forcings described in the text, the low cloud water path is varied but all other parameters are unchanged.

Colin Goldblatt^{1†} & Kevin J. Zahnle¹

¹Space Science and Astrobiology Division, NASA Ames Research Center, MS 245-3, Moffett Field, California 94035, USA.

[†]Present address: Astronomy Department and Virtual Planetary Laboratory, University of Washington, Box 351580, Seattle, Washington 98195, USA.

e-mail: cgoldbla@uw.edu

Received 1 July 2010; accepted 26 January 2011.

1. Rosing, M. T., Bird, D. K., Sleep, N. H. & Bjerrum, C. J. No climate paradox under the faint early Sun. *Nature* **464**, 744–747 (2010).

- Lin, Y. S., Heuer, V. B., Ferdelman, T. G. & Hinrichs, K.-U. Microbial conversion of inorganic carbon to dimethyl sulfide in anoxic lake sediment (Plußsee, Germany). *Biogeosciences* **7**, 2433–2444 (2010).
- Woodhouse, M. T. *et al.* Low sensitivity of cloud condensation nuclei to changes in the sea-air flux of dimethyl-sulphide. *Atmos. Chem. Phys.* **10**, 7545–7559 (2010).
- Leck, C. & Bigg, E. K. Source and evolution of the marine aerosol—a new perspective. *Geophys. Res. Lett.* **32**, L19803, doi:10.1029/2005GL023651 (2005).
- Bréon, F., Tanré, D. & Generoso, S. Aerosol effect on cloud droplet size, monitored from satellite. *Science* **295**, 834–838 (2002).
- Kump, L. R. & Pollard, D. Amplification of Cretaceous warmth by biological cloud feedbacks. *Science* **320**, 195 (2008).
- Penner, J. E. *et al.* Model intercomparison of indirect aerosol effects. *Atmos. Chem. Phys.* **6**, 3391–3405 (2006).
- Kaufman, Y. J., Koren, I., Remer, L. A., Rosenfeld, D. & Rudich, Y. The effect of smoke, dust, and pollution aerosol on shallow cloud development over the Atlantic Ocean. *Proc. Natl Acad. Sci. USA* **102**, 11207–11212 (2005).
- Trenberth, K. E., Fasullo, J. T. & Kiehl, J. T. Earth's global energy budget. *Bull. Am. Meteorol. Soc.* **90**, 311–323 (2009).
- McKay, C. P., Lorenz, R. D. & Lunine, J. I. Analytic solutions for the anti-greenhouse effect: Titan and the early Earth. *Icarus* **91**, 93–100 (1991).
- Haqq-Misra, J. D., Domagal-Goldman, S. D., Kasting, P. & Kasting, J. F. A revised, hazy methane greenhouse for the Archean Earth. *Astrobiology* **8**, 1127–1137 (2008).
- Domagal-Goldman, S. D., Kasting, J. F., Johnston, D. T. & Farquhar, J. Organic haze, glaciations and multiple sulfur isotopes in the Mid-Archean Era. *Earth Planet. Sci. Lett.* **269**, 29–40 (2008).
- Trainer, M. G. *et al.* Organic haze on Titan and the early Earth. *Proc. Natl Acad. Sci. USA* **103**, 18035–18042 (2006).
- Goldblatt, C. & Zahnle, K. J. Clouds and the faint young Sun paradox. *Clim. Past* **7**, 203–220 (2011).
- Clough, S. A. *et al.* Atmospheric radiative transfer modeling: a summary of the AER codes. *J. Quant. Spectrosc. Radiat. Transf.* **91**, 233–244 (2005).

Author Contributions C.G. and K.J.Z. discussed the article to which this note responds. C.G. performed all model runs and quantitative analysis. C.G. and K.J.Z. both contributed qualitative analysis and both contributed to writing the paper.

Competing financial interests: declared none.

doi:10.1038/nature09961

Rosing, Bird, Sleep & Bjerrum reply

REPLYING TO C. T. Reinhard & N. J. Planavsky *Nature* **474**, doi:10.1038/nature09959 (2011); N. Dauphas & J. F. Kasting *Nature* **474**, doi:10.1038/nature09960 (2011); C. Goldblatt & K. Zahnle *Nature* **474**, doi:10.1038/nature09961 (2011)

Reinhard and Planavsky¹ and Dauphas and Kasting² question whether the formation of Fe²⁺/Fe³⁺-rich sediments and the preservation of magnetite in banded iron formations (BIFs) constrains atmospheric p_{CO_2} . Goldblatt and Zahnle³ question whether reduced albedo in combination with a slightly increased greenhouse effect could compensate for the reduced luminosity of the faint young Sun. Reinhard and Planavsky¹ and Dauphas and Kasting² suggest that BIF mineralogy is solely controlled by diagenetic and metamorphic processes decoupled from ocean–atmosphere chemistry. We focus on the persistent preservation of magnetite in all Archean BIFs, and of magnetite and siderite in many BIFs⁴, and estimate the partial pressure of CO₂ during the Archean (p_{CO_2}) using the least number of assumptions concerning unconstrained parameters. Our conclusions⁵ are not undermined by any geochemical data that support high p_{CO_2} .

Diagenesis is an open-system process driven by consumption of Gibbs energy through lowering the extent of chemical disequilibrium between the sedimentary phases and surrounding environments, whether biologically mediated or not. Magnetite forms in the presence of dissimilatory iron-reducing bacteria when the pore water is supersaturated in magnetite according to standard thermodynamic principles^{6,7} and its formation is not microbially controlled^{8,9}. Isotopic data for the Kuruman BIF requires extensive transport of Fe and C between the sediment and the supernatant water by infiltration of sea water and molecular diffusion through interconnected pore space¹⁰. In modern sapropels ocean CO₂ concentration penetrates to a depth of about 15 cm in the sediment, from which point it increases more than tenfold downwards¹¹. We conclude that the CO₂ concentration in the sediment pore water would have been greater than or equal to the supernatant water. The suggestion by Dauphas and Kasting² and Reinhard and Planavsky¹ that diagenetic siderite formation was a closed-system process is not supported by the studies they quote, which call upon extensive communication with the supernatant water^{10,12}.

Formation of BIFs depends not only on mechanisms for precipitation of solid Fe compounds but also on the transport efficiency of dissolved iron. The saturation constraints on p_{CO_2} and p_{H_2} imposed by siderite and magnetite not only pertain to the BIF, but also to the water body supplying dissolved Fe to sedimentary basins, and thus couple

directly to the atmosphere. We note that for our estimated range of p_{CO_2} the solubility of Fe in sea water is 1–3 μM , which is well within the range necessary for BIF formation¹³, while at the increased p_{CO_2} favoured by Reinhard and Planavsky¹ and Dauphas and Kasting² the maximum Fe concentrations are between two and four orders of magnitude lower and within the range of modern ocean water concentrations that preclude BIF formation. (We note that the 1–3 μM range covers the bare Fe²⁺ ion corrected for activity in sea water at siderite saturation with equilibrium constant for the reaction $\text{Fe}^{2+} + \text{CO}_{2,\text{g}} + \text{H}_2\text{O} = \text{FeCO}_3 + 2\text{H}^+$ calculated in SUPCRT (see ref. 5) and total Fe calculated using EQ3-6 (ref. 14) for siderite saturation in modern sea water, corrected to anoxic conditions and assuming pH = 8.25 buffered by basalt in accordance with ref. 15. See ref. 5 for thermodynamic conventions and data sources.)

Reinhard and Planavsky¹ suggest that siderite and haematite can coexist at 0.1 nM H₂ and p_{CO_2} in the range 30–100 PAL at 25 °C–35 °C, in accord with minimum H_{2(aq)} metabolic constraints of dissimilatory iron-reducing bacteria and phase relations in Fig. 1 of ref. 5. This logic precludes the presence of magnetite in BIFs. Neglecting observed magnetite parageneses in BIFs and selecting a value of H_{2(aq)} far below the stability field of magnetite does not provide a viable geochemical constraint. Dissimilatory iron-reducing bacteria metabolism may have occurred in BIF sediments during the Archean, but we are not aware of geochemical evidence suggesting that the metabolic activity of dissimilatory iron-reducing bacteria in BIF sediments controlled the composition of the Earth's fluid envelopes. On the contrary, the presence of magnetite and the mere formation of BIFs indicate that ambient ocean water had a molarity of H₂ considerably greater than 10^{−10} M, conditions under which haematite is the only stable Fe oxide and Fe would be practically insoluble in ocean water, precluding aqueous iron transport sufficient to sustain BIF sedimentation.

It has never been controversial that there is a large uncertainty about Archean cloud forcing. Our hypothesis reiterates this specifically for an Earth with no continents, no continental dust, no land-plant-generated secondary organic aerosols and only a few other biogenic aerosols from a low-productivity ocean (0.1 to 0.25 times the present value^{16,17}). The cloud physics are affected by an aerosol optical

depth that probably showed much lower global values than today and should be evaluated using proper general circulation models. Goldblatt and Zahnle³ miss the point when focusing on dimethyl sulphide from a sulphur-poor Archean ocean and sea-salt aerosols, which seem to contribute only a few per cent of the free troposphere cloud condensation nuclei¹⁸.

The net cloud radiative effect of our model was (-18 W m^{-2}) relative to -20 W m^{-2} from satellite estimates. Our calculations used¹⁹ an effective radius of cloud droplets of $r_e = 17 \mu\text{m}$ (green solid line in Fig. 2d of ref. 5) (we note that although the legend mentions $20 \mu\text{m}$, this was erroneous, and should have stated $17 \mu\text{m}$; also, we showed the line for $30 \mu\text{m}$ to show no added effect at larger droplet sizes). Goldblatt and Zahnle's later published calculations²⁰ use more realistic clouds, which are not very different from ours. Our model values versus theirs¹⁹ were: r_e increase to $16.5 \mu\text{m}$, 7 W m^{-2} versus 7 W m^{-2} ; enhanced rain-out, 15 W m^{-2} versus $4\text{--}15 \text{ W m}^{-2}$; aqua planet, 6 W m^{-2} versus 5 W m^{-2} ; $1,000 \text{ p.p.m.v. CO}_2$, 5 W m^{-2} versus 6 W m^{-2} . Our methane contribution was a bit high (9 W m^{-2} ; ref. 17's was 7 W m^{-2}) and should probably be adjusted down to $0.3 \times p_{\text{CO}_2}$. Rain-out of low clouds (we made no change to high clouds) was calibrated to general circulation model results²¹, and an estimated forcing three times the r_e forcing¹⁸. Our calculations showed⁵, as subsequently stated by Goldblatt and Zahnle²⁰, that "...precipitation feedback is of first order importance and must be treated carefully in any model addressing the climatic effect of changed particle size."

We maintain that formation and preservation of magnetite-rich BIF could not take place under a high- p_{CO_2} atmosphere. Surface albedo and cloud changes would have significantly decreased the Earth's albedo and reduced the greenhouse gas concentrations needed for clement conditions on the Earth. We do not claim to have provided a final and unique solution to the 'faint young Sun paradox', but rather demonstrate that the presence of liquid water on the young Earth can be explained within uncertainty by constrainable parameters and is thus not so paradoxical.

Minik T. Rosing^{1,2,3}, Dennis K. Bird^{1,2}, Norman H. Sleep⁴ & Christian J. Bjerrum^{1,5}

¹Nordic Center for Earth Evolution, Øster Voldgade 5-7, DK-1350 København K., Denmark.

e-mail: minik@snm.ku.dk

²Department of Geological and Environmental Sciences, Stanford University, Stanford, California 94305, USA.

³Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, DK-1350 København K., Denmark.

⁴Department of Geophysics, Stanford University, Stanford, California 94305, USA.

⁵Department of Geography and Geology, University of Copenhagen, Øster Voldgade 10, DK-1350 København K., Denmark.

- Reinhard, C. T. & Planavsky, N. J. Mineralogical constraints on Precambrian p_{CO_2} . *Nature* **474**, doi:10.1038/nature09959 (this issue).
- Dauphas, N. & Kasting, J. F. Low p_{CO_2} in the pore water, not in the Archean air. *Nature* **474**, doi:10.1038/nature09960 (this issue).
- Goldblatt, C. & Zahnle, K. Faint young Sun paradox remains. *Nature* **474**, doi:10.1038/nature09961 (this issue).
- Klein, C. Some Precambrian banded iron-formations (BIFs) from around the world: their age, geologic setting, mineralogy, metamorphism, geochemistry, and origin. *Am. Mineral.* **90**, 1473–1499 (2005).
- Rosing, M. T., Bird, D. K., Sleep, N. H. & Bjerrum, C. J. No climate paradox under the faint early Sun. *Nature* **464**, 744–747 (2010).
- Kocar, B. D. & Fendorf, S. Thermodynamic constraints on reductive reactions influencing the biogeochemistry of arsenic in soils and sediments. *Environ. Sci. Technol.* **43**, 4871–4877 (2009).
- Hansel, C. M. *et al.* Secondary mineralization pathways induced by dissimilatory iron reduction of ferrihydrite under advective flow. *Geochim. Cosmochim. Acta* **67**, 2977–2992 (2003).
- Behrends, T. & Van Cappellen, P. Transformation of hematite into magnetite during dissimilatory iron reduction—conditions and mechanisms. *Geomicrobiol. J.* **24**, 403–416 (2007).
- Lovley, D. R. & Goodwin, S. Hydrogen concentrations as an indicator of the predominant terminal electron-accepting reactions in aquatic sediments. *Geochim. Cosmochim. Acta* **52**, 2993–3003 (1988).
- Heimann, A. *et al.* Fe, C, and O isotope compositions of banded iron formation carbonates demonstrate a major role for dissimilatory iron reduction in ~2.5 Ga marine environments. *Earth Planet. Sci. Lett.* **294**, 8–18 (2010).
- Boudreau, B. P., Canfield, D. E. & Mucci, A. Early diagenesis in a marine sapropel, Mangrove Lake, Bermuda. *Limnol. Oceanogr.* **37**, 1738–1753 (1992).
- Becker, R. H. & Clayton, R. N. Carbon isotopic evidence for origin of a banded iron-formation in Western Australia. *Geochim. Cosmochim. Acta* **36**, 577–595 (1972).
- Konhauser, K. O., Newman, D. K. & Kappler, A. The potential significance of microbial Fe(III) reduction during deposition of Precambrian banded iron formations. *Geobiology* **3**, 167–177 (2005).
- Wolery, T. J. *EQ3nr. A Computer Program for Geochemical Aqueous Speciation Solubility Calculations: Theoretical Manual, User's Guide and Related Documentation*. Version 7.0 (Lawrence Livermore National Laboratory, 1991).
- Friend, C. R. L., Nutman, A. P., Bennett, V. C. & Norman, M. D. Seawater-like trace element signatures (REE+Y) of Eoarchaeon chemical sedimentary rocks from southern West Greenland, and their corruption during high-grade metamorphism. *Contrib. Mineral. Petrol.* **155**, 229–246 (2008).
- Bjerrum, C. J. & Canfield, D. E. Ocean productivity before about 1.9 Gyr ago limited by phosphorus adsorption onto iron oxides. *Nature* **417**, 159–162 (2002).
- Canfield, D. E., Rosing, M. T. & Bjerrum, C. Early anaerobic metabolisms. *Phil. Trans. R. Soc. Lond. B* **361**, 1819–1834 (2006).
- Andreae, M. O. & Rosenfeld, D. Aerosol-cloud-precipitation interactions. Part 1. The nature and sources of cloud-active aerosols. *Earth Sci. Rev.* **89**, 13–41 (2008).
- Kaufman, Y. J., Koren, I., Remer, L. A., Rosenfeld, D. & Rudich, Y. The effect of smoke, dust, and pollution aerosol on shallow cloud development over the Atlantic Ocean. *Proc. Natl Acad. Sci. USA* **102**, 11207–11212 (2005).
- Goldblatt, C. & Zahnle, K. J. Clouds and the faint young Sun paradox. *Clim. Past Discuss.* **6**, 1163–1207 (2010).
- Kump, L. R. & Pollard, D. Amplification of Cretaceous warmth by biological cloud feedbacks. *Science* **320**, 195 (2008).

Author Contributions All authors contributed equally to this Reply.

doi:10.1038/nature09962

Regulation of angiogenesis by a non-canonical Wnt–Flt1 pathway in myeloid cells

James A. Stefater III^{1,2}, Ian Lewkowich³, Sujata Rao^{1,2}, Giovanni Mariggi⁴, April C. Carpenter^{1,2}, Adam R. Burr⁵, Jieqing Fan^{1,2}, Rieko Ajima⁶, Jeffery D. Molkentin^{5,7}, Bart O. Williams⁸, Marsha Wills-Karp³, Jeffrey W. Pollard⁹, Terry Yamaguchi⁶, Napoleone Ferrara¹⁰, Holger Gerhardt^{4,11} & Richard A. Lang^{1,2}

Myeloid cells are a feature of most tissues. Here we show that during development, retinal myeloid cells (RMCs) produce Wnt ligands to regulate blood vessel branching. In the mouse retina, where angiogenesis occurs postnatally¹, somatic deletion in RMCs of the Wnt ligand transporter *Wntless*^{2,3} results in increased angiogenesis in the deeper layers. We also show that mutation of *Wnt5a* and *Wnt11* results in increased angiogenesis and that these ligands elicit RMC responses via a non-canonical Wnt pathway. Using cultured myeloid-like cells and RMC somatic deletion of *Flt1*, we show that an effector of Wnt-dependent suppression of angiogenesis by RMCs is Flt1, a naturally occurring inhibitor of vascular endothelial growth factor (VEGF)^{4–6}. These findings indicate that resident myeloid cells can use a non-canonical, Wnt–Flt1 pathway to suppress angiogenic branching.

Myeloid cells have a wide array of biological activities that include immune activation, arteriogenesis⁷, and regulation of salt balance and blood pressure⁸. Myeloid cells also regulate vascularity. Tumour-associated macrophages influence the growth of blood vessels⁹ in part because they are a source of VEGFA¹⁰. Myeloid cells also promote angiogenic branching¹¹ and anastomosis¹². Depending on the context, myeloid cells can be either anti-angiogenic¹³ or pro-angiogenic¹⁴. Here we show that RMCs suppress retinal angiogenesis via a Wnt–Flt1 pathway (Supplementary Fig. 1).

Retinal angiogenesis begins on the day of birth in the mouse with the formation of a superficial vascular plexus (Fig. 1a) that lies within the ganglion cell layer¹⁵. After formation of this superficial plexus by postpartum day 7 (P7), angiogenic sprouts descend vertically through the retinal layers from P8 to P14 (Fig. 1a). At the outer edge of the inner nuclear layer (INL) the vertical angiogenic sprouts turn and simultaneously branch to form a deep vascular plexus (Fig. 1a). Using antibodies to the vascular endothelial cell (VEC) marker endomucin and to the green fluorescent protein (GFP) of the *c-fms-EGFP* (also known as *Tg(Csf1r-EGFP)1Hume*) transgene that marks RMCs, we show that myeloid cells have a unique spatial relationship with angiogenic tip cells. At the point of turning and branching in the outer INL, myeloid cells and angiogenic sprouts are in close contact (Supplementary Fig. 2a). This was confirmed by labelling with isolectin and F4/80 (F4/80 is also known as *Emr1*; Supplementary Fig. 2b). We then took advantage of high-intensity isolectin labelling of VECs and RMCs and performed a three-dimensional reconstruction with false colouring to illustrate the overall topology of the angiogenic tip cell–RMC interaction (Fig. 1b). This showed close contact between the two cell types throughout turn-and-branch angiogenesis in the deep retinal layer. Furthermore, after turning, angiogenic tip cells extend within the plane of the deep retinal layer and remain RMC-associated (Fig. 1c). In further defining RMCs,

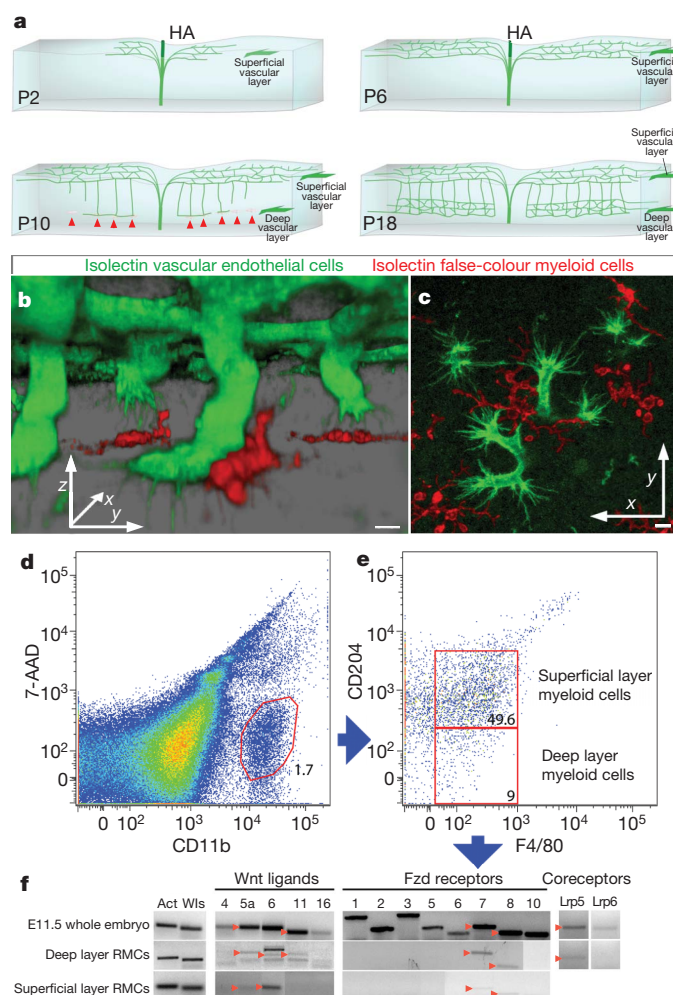


Figure 1 | RMCs interact with VECs and express Wnt components. **a**, Schematic of retina at postnatal days (P) 2, 6, 10, and 18. RMCs interacting with descending vertical sprouts are labelled with red arrowheads. Adapted from ref. 1. **b**, Isolectin-labelled three-dimensional reconstruction of vertical angiogenic sprouts (green) and RMCs (false-colour red). **c**, As in **b** but a two-dimensional image in the deep vascular layer. Scale bars, 5 μ m. **d**, **e**, Flow cytometry of deep layer RMCs based on surface markers. 7-AAD, 7-aminoactinomycin D. **f**, PCR for Wnt pathway components on flow-sorted RMCs. Red arrowheads indicate expected sizes. Act, β -actin.

¹The Visual Systems Group, Divisions of Pediatric Ophthalmology and Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA. ²Department of Ophthalmology, University of Cincinnati, Cincinnati, Ohio 45229, USA. ³Division of Immunobiology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA. ⁴Vascular Biology Laboratory, London Research Institute, Cancer Research UK, London WC2 3PX, UK. ⁵Division of Molecular Cardiovascular Biology, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Ohio 45229, USA. ⁶Cancer and Developmental Biology Laboratory, National Cancer Institute, Frederick, Maryland 21701, USA. ⁷Howard Hughes Medical Institute, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Ohio 45229, USA. ⁸Center for Skeletal Disease Research, Van Andel Research Institute, 333 Bostwick NE, Grand Rapids, Michigan 49503, USA. ⁹Albert Einstein College of Medicine of Yeshiva University, Jack and Pearl Resnick Campus, 1300 Morris Park Avenue, Bronx, New York 10461, USA. ¹⁰Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. ¹¹Consultant Group Leader, Vascular Patterning Laboratory, Vesalius Research Center, VIB, Campus Gasthuisberg, B-3000 Leuven, Belgium.

we showed that CD204 (also known as *Msrl*) was unique to RMCs with amoeboid morphology in the superficial retinal layer (Supplementary Fig. 2c, d). By contrast, deep retinal layer RMCs with extended morphology were CD204-negative (Supplementary Fig. 2c, e). Both layers of RMCs expressed *Iba1* (ref. 16) but at different levels (Supplementary Fig. 2c–e). This information allowed us to sort distinct populations of superficial (CD11b+, F4/80+, CD204+) and deep layer (CD11b+, F4/80+, CD204-) RMCs by flow cytometry (Fig. 1d, e).

On the basis of previous work showing vascular regulation by myeloid Wnt ligands¹⁷ we proposed that RMCs might use Wnt ligands to regulate retinal angiogenesis. First, we examined the expression of Wnt ligands and receptors in superficial and deep RMCs isolated by flow cytometry. PCR with reverse transcription (RT-PCR) analysis of the deep RMC population showed expression of *Wnt5a*, *Wnt6* and *Wnt11*, *Fzd7* and *Fzd8* as well as the co-receptor *Lrp5* (Fig. 1f). With the exception of *Wnt5b*, which was inconsistent, no other Wnt ligand was detected in deep RMCs. Superficial RMCs expressed similar Wnt and *Fzd* proteins (Fig. 1f), but also expressed *Wnt2b*, 3 and 3a (data not shown).

The challenge of genetic analysis when RMCs express many Wnt ligands was addressed by the generation of a *loxP*-flanked conditional

allele for the essential Wnt ligand transporter *Wls*^{2,3}. Both superficial and deep RMCs expressed *Wls* (Fig. 1f). *Wls* was deleted using the myeloid cre driver *cfms-icre*¹⁸, which we confirmed was functional in RMCs (Supplementary Fig. 3a–c). To analyse retinal vasculature, we imaged superficial and deep retinal vessels at P18 (Fig. 2a). Quantification of vessel branch points showed that *cfms-icre* alone had no effect (Fig. 3b). Furthermore, compared with control *Wls*^{fl/+} mice, *Wls*^{fl/+}; *cfms-icre* animals had a normal superficial vascular plexus (Fig. 2a, c). By contrast, the deep vascular layer (Fig. 2a, c) showed an overgrowth. Interestingly, no further enhancement of vascular overgrowth was apparent when the myeloid *Wls* deletion was homozygous as in *Wls*^{-/-}; *cfms-icre* mice (data not shown). Because myeloid cells are positioned below descending vessels at P10, we assessed vessel branching at the base of these sprouts. *Wls*^{fl/+}; *cfms-icre* mice showed reduced simple turning (no branches) and single branching, but significantly more multi-branch events (Fig. 2d). As a weighted mean, the branch index was 2.0 in control and 3.2 in the mutant ($P < 0.0001$).

The higher vascular density of the deep layer in *Wls*^{fl/+}; *cfms-icre* mice could reflect enhanced angiogenesis or defective remodelling. To

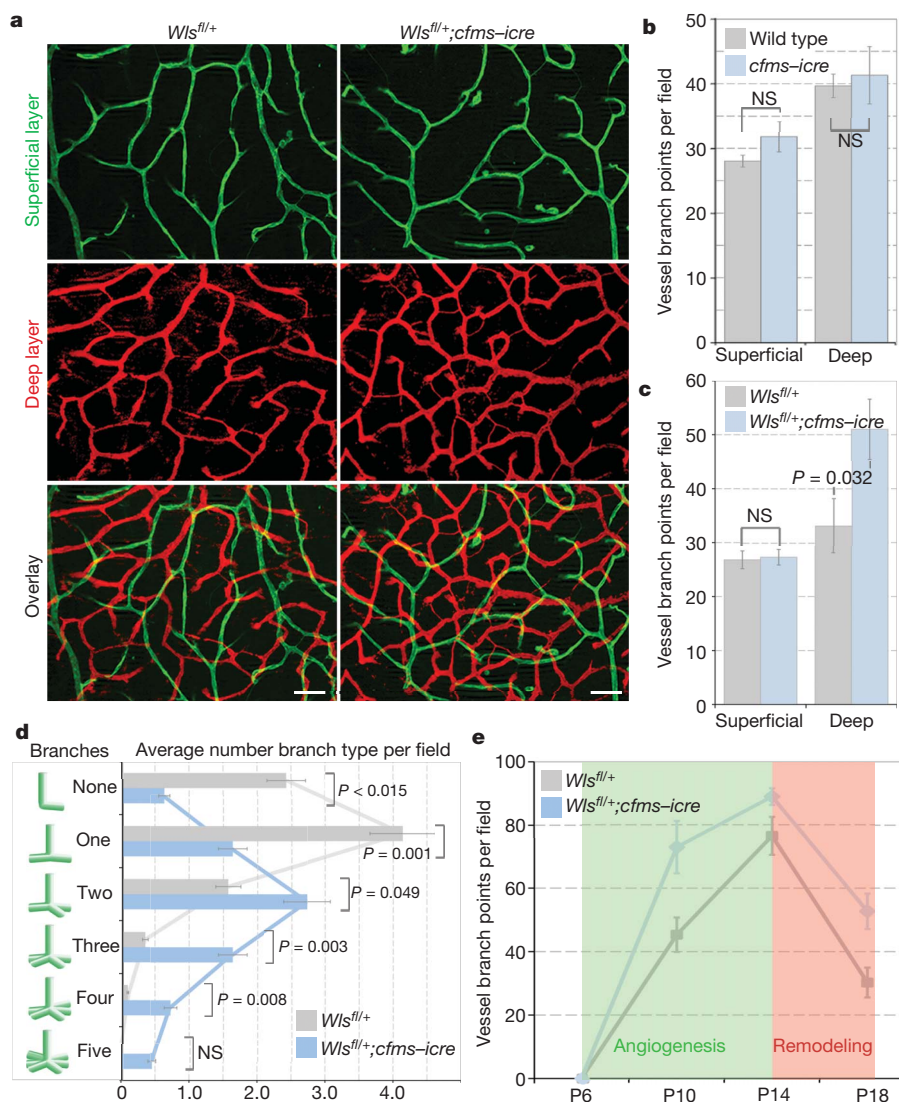


Figure 2 | RMC *Wls* is required for suppression of deep angiogenic branching. **a**, Isolectin labelling of superficial and deep retinal vasculature in *Wls*^{fl/+} and *Wls*^{fl/+}; *cfms-icre* mice. Scale bars, 50 μ m. **b**, **c**, P18 vessel branch points in labelled genotypes. $n = 4$ (b), $n = 8$ (c). **d**, Branches emanating from the base of vertical sprouts in the P10 deep vascular layer. $n = 8$. **b–d** used

Student's *t*-test (two-tailed). **e**, Time-course of deep layer branches in indicated genotypes. Shading shows when angiogenesis (green) and remodelling (red) predominate. One-way ANOVA showed $P = 0.0021$. $n = 4$ for each point. Error bars are s.e.m., NS, not significant.

assess this, we counted deep layer branch points in control ($Wls^{fl/+}$) and experimental ($Wls^{fl/+}; cfms-icre$) mice over a time-course from P6 to P18 (Fig. 2e). Between P14 and P18 when remodelling predominates (Fig. 2e, red zone) control and mutant graph slopes were nearly identical. By contrast, between P6 and P14 when vessel growth predominates (Fig. 2e, green zone), the slope of the graph is greater in the mutant. There was no difference between control and mutant animals in the number of vertical sprouts connecting superficial and deep layers (Supplementary Fig. 3d). Combined, the data in Fig. 2 are consistent with a model in which myeloid Wnt ligands suppress branch formation as angiogenic sprouts make contact with deep layer RMCs (Fig. 1b, c). These data were corroborated by the vascular overgrowth phenotype of *Wnt5a* (ref. 19) and *Wnt11* (ref. 20) heterozygotes (Supplementary Figs 3e, f and 4).

Wnt5a and *Wnt11* are most often associated with non-canonical Wnt signalling²¹. We thus investigated the possibility that Wnt-dependent suppression of deep retinal angiogenesis was a non-canonical response. *Wnt3a*, but not *Wnt5a*, elicited a canonical response in SuperTopflash (STF) reporter cells (Supplementary Fig. 5a). Some non-canonical Wnt pathways activate a Ca^{2+} flux²¹. To determine whether myeloid Wnt proteins elicited a Ca^{2+} flux, we added *Wnt5a* to myeloid-like RAW264.7 cells and measured Ca^{2+} -dependent Fura-2 dye emission. Recombinant *Wnt5a* increased intracellular Ca^{2+} compared to untreated controls (Fig. 3a and Supplementary Fig. 5b, c). Furthermore, to assess the requirement for Wls in Wnt ligand secretion, we exposed RAW264.7 cells to medium from $Wls^{fl/fl}$ mouse embryonic fibroblasts (MEFs) transfected with *Wnt5a* or *Wnt5a* and *cre* recombinase plasmids. RAW264.7 cells had increased Ca^{2+} flux in response to *Wnt5a*-transfected MEF medium relative to *Wnt5a*- and *cre*-transfected medium (Fig. 3b). This validates the role of Wls in secretion of *Wnt5a*. Combined, these data show that one myeloid Wnt, *Wnt5a*, does not stimulate a canonical Wnt response, but can elicit a Ca^{2+} flux characteristic of some non-canonical Wnt responses.

In recent work, it has been shown that *Wnt5a* loss-of-function can rescue the defects associated with deletion of the canonical Wnt pathway co-receptors *Lrp5* and *Lrp6* (ref. 22). This finding has suggested that *Lrp5/6* deletion actually represents a non-canonical pathway gain-of-function and that in deleting the non-canonical ligand *Wnt5a*, there is a re-balancing of pathway activity. This hypothesis is supported by biochemical analysis showing that *Wnt5a* can bind *Lrp6*, but does not elicit the phosphorylation required for canonical signalling²². Canonical, non-canonical reciprocal pathway inhibition has also been demonstrated²³. These findings argue that a deletion of *Lrp5/6* can define whether a Wnt signalling pathway is canonical or non-canonical. If it is canonical, an *Lrp5/6* deletion will give the same phenotype as ligand deletion. By contrast, if the signalling pathway is non-canonical, the consequence of *Lrp5/6* deletion would be opposite to that of ligand mutation. Thus, to determine whether the RMC response was canonical or non-canonical, we generated a *cfms-icre* somatic mutant of the *Lrp5* coreceptor that is expressed in RMCs (Fig. 1f). Because the result was significantly diminished deep vascular layer density in somatic homozygotes (Fig. 3c–f), a response opposite to ligand deletion, this provides *in vivo* evidence that the Wnt response is non-canonical. Even though the superficial vascular layer was slightly deficient in somatic homozygotes (Fig. 3e) this was not the reason for reduced density in the deep vascular layer as the descending sprout number was unchanged (Fig. 3f).

The VEGF receptor *Flt1* (also known as *VEGFR1*) can suppress angiogenesis because it has limited signalling capacity, a higher affinity for VEGF than *Flk1* (also known as *VEGFR2*) and can sequester *VEGF*^{5,6}. Alternative splicing produces both membrane-tethered and soluble forms⁵. *Flt1* is known to contribute to corneal avascularity⁴ as well as the selection of angiogenic tip cells when expressed regionally in an existing vessel²⁴. *Flt1* is also known to be expressed in some myeloid populations⁶. Because angiogenesis in the deep retinal layers is VEGF-dependent²⁵, *Flt1* was a good candidate to mediate Wnt-dependent angiogenic suppression by deep RMCs.

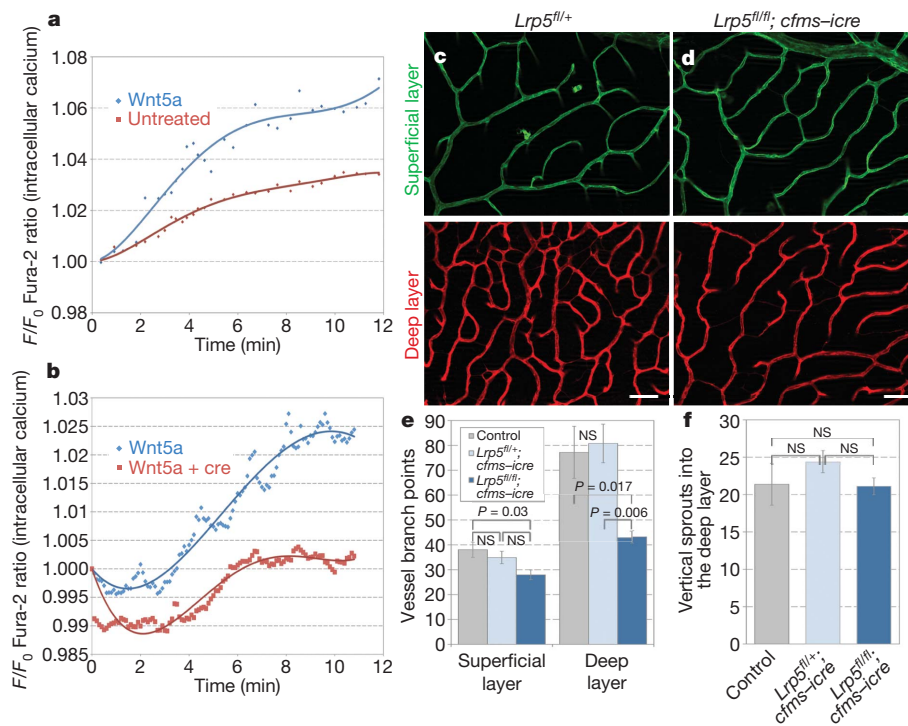


Figure 3 | Angiogenic suppression by RMCs is a non-canonical Wnt response. **a, b**, Intracellular Ca^{2+} in RAW264.7 cells treated with *Wnt5a* (**a**) or supernatant from $Wls^{fl/fl}$ MEFs expressing *Wnt5a* or *Wnt5a* and *cre* (**b**). One Way ANOVA showed $P \leq 0.0001$ for both. **c, d**, Isolectin labelling of superficial and deep retinal vasculature in $Lrp5^{fl/+}$ (**c**) and $Lrp5^{fl/fl}; cfms-icre$ (**d**) mice.

50 μ m scale bars. **e**, P18 vessel branch points in labelled genotypes. **f**, Vertical vessels connecting to the deep layer in labelled genotypes. **e, f**, Used one-way ANOVA with Tukey's post-hoc. $n \geq 8$ per genotype. Errors are s.e.m., NS, not significant.

We thus determined whether *Flt1* was expressed in P12 RMCs by performing end-point RT-PCR on flow-sorted cells using primers that detected both transcripts. This showed that deep RMCs expressed *Flt1* (Fig. 4a). By contrast, superficial RMCs did not (Fig. 4a). We also showed that Iba1-labelled RMCs positioned at the outer edge of the INL labelled with anti *Flt1* antibodies (Fig. 4b). We then generated an RMC *Flt1* loss-of-function using *cfms-icre* and the *Flt1^{fl/+}* conditional allele²⁶. Quantification of retinal vascular branch points in *Flt1^{fl/+}*; *cfms-icre* mice at P18 showed that the superficial layer was unaffected but that the deep layer showed an increase in density (Fig. 4c, d). As with *Wls* somatic mutants, conditional homozygosity for *Flt1* (*Flt1^{fl/fl}*; *cfms-icre*) did not give a further significant increase in vascular density (data not shown). Furthermore, the number of deep RMCs was unchanged (Supplementary Fig. 2g).

These data showed that conditional deletion of *Flt1* produced an enhancement of angiogenesis similar to that observed with conditional

deletion of *Wls*. One implication was that *Wls* and *Flt1* might function in the same angiogenesis suppression pathway. To test the possibility that *Flt1* might be regulated by Wnt ligands, we used myeloid-like RAW264.7 cells that, like RMCs, express *Fzd7* and *Fzd8* (Fig. 4e). When stimulated with the 'canonical' ligand Wnt3a, RAW264.7 did not change the level of *Flt1* transcript according to quantitative PCR (Fig. 4f). However, stimulation with Wnt5a produced a threefold increase in the level of both *Flt1* isoform transcripts (Fig. 4f). Furthermore, an enzyme-linked immunosorbent assay (ELISA) showed that there was an increased level of soluble *Flt1* in conditioned media from Wnt5a-stimulated RAW264.7 cells (Fig. 4g). Wnt3a stimulation again had no effect (Fig. 4g).

As a stringent test of the possibility that Wnt ligands stimulated expression of *Flt1* in deep RMCs, we flow-sorted the CD11b⁺, F4/80⁺, CD204⁺ population from both control (*Wls^{fl/+}*) and experimental (*Wls^{fl/+}*; *cfms-icre*) mice at P12 and performed quantitative PCR for

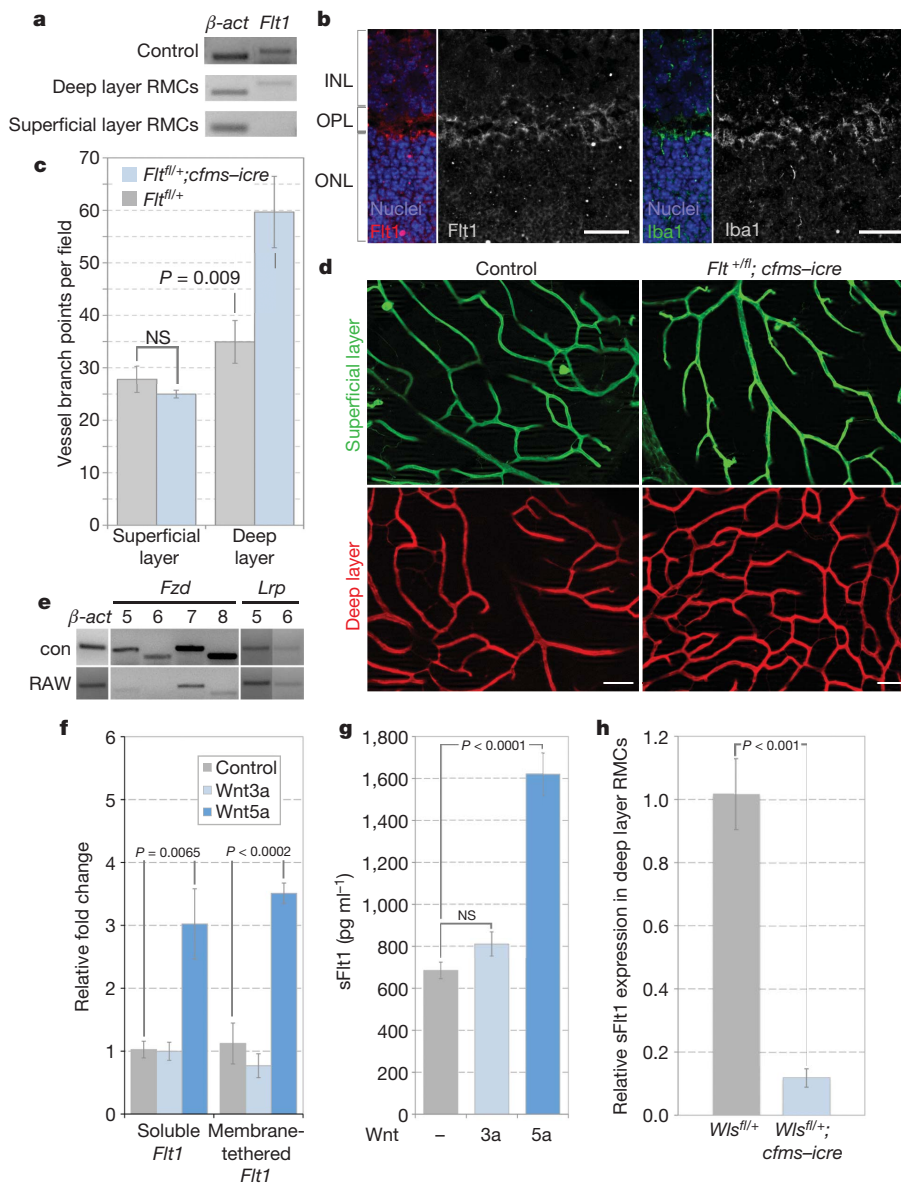


Figure 4 | *Flt1* expression in myeloid cells is regulated by a Wnt pathway. **a**, PCR for *Flt1* transcript in flow-sorted RMCs. β -actin, β -actin transcript. **b**, P14 retinal sections labelled for *Flt1* and Iba1. Scale bars, 25 μ m. **c**, P18 vessel branch points in the superficial and deep vasculature in *Flt1^{fl/+}* and *Flt1^{fl/+}*; *cfms-icre* mice ($n = 7$, Student's *t*-test). $n = 7$. **d**, Isolectin labelling of the retinal vasculature in labelled genotypes. Scale bars, 50 μ m. **e**, PCR for indicated

transcripts in E11.5 whole embryo (con) and RAW264.7 cells. **f**, **g**, quantitative PCR (f) for soluble *Flt1* and membrane-tethered *Flt1* in Wnt-treated RAW264.7 cells and ELISA (g) for soluble *Flt1* on medium from Wnt-treated RAW264.7 cells ($n = 4$, one-way ANOVA). **h**, Quantitative PCR for soluble *Flt1* in flow-sorted deep RMCs from *Wls^{fl/+}* and *Wls^{fl/+}*; *cfms-icre* mice ($n = 4$, Student's *t*-test). Errors are s.e.m.; NS, not significant.

Flt1 transcripts. We were never able to amplify the membrane-tethered *Flt1* transcript but showed that there was a 90% reduction in transcript level for soluble *Flt1* (Fig. 4h). This was consistent with the maximal phenotype observed in the heterozygous conditional *Wls* mutants. Although a heterozygous phenotype is not unusual, it is perhaps surprising that conditional homozygosity gave no further significant change. This kind of response might be explained by a low signalling sensitivity threshold for the Wnt–*Flt1* pathway.

The data we describe show that RMCs can modulate angiogenic responses directly by producing the VEGF inhibitory receptor *Flt1*. Unexpectedly, we also show that the production of *Flt1* depends on myeloid non-canonical Wnt ligands. In this setting, myeloid Wnt ligands might function in autocrine stimulation as has been documented for cultured macrophages²⁷ or might operate via a more complex pathway involving another cell type as an intermediate. The Wnt–*Flt1* response represents a new pathway for the regulation of VEGF-stimulated angiogenesis. The targeting of appropriate *Fzd* receptors or other Wnt–*Flt1* pathway components may offer new opportunities to modulate the production of *Flt1* and thus, the VEGF-stimulated angiogenic response. Because macrophage-related cells are ubiquitous and highly mobile, it is possible that the Wnt–*Flt1* pathway will be a general means to suppress VEGF locally and thus to restrain vascular responses. In future studies it will be interesting to determine, for example, if this pathway is active in the suppression of wound angiogenesis by macrophages¹³ or is inactivated in macrophage-dependent tumour angiogenesis²⁸.

Additionally, our observations are consistent with a string of recent papers describing the role of myeloid cells beyond their well-documented functions in innate immunity^{7,8,17,29}. These findings build on an idea articulated nearly a century ago³⁰. It was suggested then that phagocytes originally evolved to regulate developmental processes, and that their immune functions were a later evolutionary adaptation. Macrophages were proposed to be the “policemen” of multi-cellular organisms, and that they could establish “harmony from chaos”. Here we show, in the setting of the retina, that myeloid cells do just that by fine-tuning vascular density and directing vascular traffic.

METHODS SUMMARY

We prepared and stained retinas as reported previously¹. To isolate RMCs, we digested retinas, pre-enriched with CD11b beads, and sorted for surface markers with the FACSria II. To obtain conditioned medium from MEFs, we performed transient transfections with *Wnt5a*, *Thy1.1* and *cre* plasmids, sorted *Thy1.1*-positive MEFs, and re-plated transfected cells. We performed calcium imaging on RAW264.7 cells loaded in Ringer's solution with 5 μ M Fura-2 AM and imaged at 510 nm at 1 Hz after excitation at 340 nm and 380 nm. All animal experiments were performed in accordance with IACUC-approved guidelines and regulations.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 July 2010; accepted 30 March 2011.

Published online 29 May 2011.

- Gerhardt, H. *et al.* VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. *J. Cell Biol.* **161**, 1163–1177 (2003).
- Ching, W. & Nusse, R. A dedicated Wnt secretion factor. *Cell* **125**, 432–433 (2006).
- Carpenter, A. C., Rao, S., Wells, J. M., Campbell, K. & Lang, R. A. Generation of mice with a conditional null allele for *Wntless*. *Genesis* **48**, 554–558 (2010).
- Ambati, B. K. *et al.* Corneal avascularity is due to soluble VEGF receptor-1. *Nature* **443**, 993–997 (2006).
- Kendall, R. L. & Thomas, K. A. Inhibition of vascular endothelial cell growth factor activity by an endogenously encoded soluble receptor. *Proc. Natl Acad. Sci. USA* **90**, 10705–10709 (1993).
- Shibuya, M. Structure and dual function of vascular endothelial growth factor receptor-1 (*Flt-1*). *Int. J. Biochem. Cell Biol.* **33**, 409–420 (2001).

- Pipp, F. *et al.* VEGFR-1-selective VEGF homologue PIGF is arteriogenic: evidence for a monocyte-mediated mechanism. *Circ. Res.* **92**, 378–385 (2003).
- Machnik, A. *et al.* Macrophages regulate salt-dependent volume and blood pressure by a vascular endothelial growth factor-C-dependent buffering mechanism. *Nature Med.* **15**, 545–552 (2009).
- Lin, E. Y. & Pollard, J. W. Tumor-associated macrophages press the angiogenic switch in breast cancer. *Cancer Res.* **67**, 5064–5066 (2007).
- Stockmann, C. *et al.* Deletion of vascular endothelial growth factor in myeloid cells accelerates tumorigenesis. *Nature* **456**, 814–818 (2008).
- Kubota, Y. *et al.* M-CSF inhibition selectively targets pathological angiogenesis and lymphangiogenesis. *J. Exp. Med.* **206**, 1089–1102 (2009).
- Fantin, A. *et al.* Tissue macrophages act as cellular chaperones for vascular anastomosis downstream of VEGF-mediated endothelial tip cell induction. *Blood* **116**, 829–840 (2010).
- Martin, P. *et al.* Wound healing in the PU.1 null mouse—tissue repair is not dependent on inflammatory cells. *Curr. Biol.* **13**, 1122–1128 (2003).
- Grunewald, M. *et al.* VEGF-induced adult neovascularization: recruitment, retention, and role of accessory cells. *Cell* **124**, 175–189 (2006).
- Saint-Geniez, M. & D'Amore, P. A. Development and pathology of the hyaloid, choroidal and retinal vasculature. *Int. J. Dev. Biol.* **48**, 1045–1058 (2004).
- Mendes-Jorge, L. *et al.* Scavenger function of resident autofluorescent perivascular macrophages and their contribution to the maintenance of the blood-retinal barrier. *Invest. Ophthalmol. Vis. Sci.* **50**, 5997–6005 (2009).
- Lobov, I. B. *et al.* WNT7b mediates macrophage-induced programmed cell death in patterning of the vasculature. *Nature* **437**, 417–421 (2005).
- Deng, L. *et al.* A novel mouse model of inflammatory bowel disease links mammalian target of rapamycin-dependent hyperproliferation of colonic epithelium to inflammation-associated tumorigenesis. *Am. J. Pathol.* **176**, 952–967 (2010).
- Yamaguchi, T. P., Bradley, A., McMahon, A. P. & Jones, S. A. Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. *Development* **126**, 1211–1223 (1999).
- Majumdar, A., Vainio, S., Kispert, A., McMahon, J. & McMahon, A. P. Wnt11 and Ret/Gdnf pathways cooperate in regulating ureteric branching during metanephric kidney development. *Development* **130**, 3175–3185 (2003).
- Seifert, J. R. & Mlodzik, M. Frizzled/PCP signalling: a conserved mechanism regulating cell polarity and directed motility. *Nature Rev. Genet.* **8**, 126–138 (2007).
- Bryja, V. *et al.* The extracellular domain of Lrp5/6 inhibits noncanonical Wnt signaling *in vivo*. *Mol. Biol. Cell* **20**, 924–936 (2009).
- Grumolato, L. *et al.* Canonical and noncanonical Wnts use a common mechanism to activate completely unrelated coreceptors. *Genes Dev.* **24**, 2517–2530 (2010).
- Chappell, J. C., Taylor, S. M., Ferrara, N. & Bautch, V. L. Local guidance of emerging vessel sprouts requires soluble Flt-1. *Dev. Cell* **17**, 377–386 (2009).
- Haigh, J. J. *et al.* Cortical and retinal defects caused by dosage-dependent reductions in VEGF-A paracrine signaling. *Dev. Biol.* **262**, 225–241 (2003).
- Lichtenberger, B. M. *et al.* Autocrine VEGF signaling synergizes with EGFR in tumor cells to promote epithelial cancer development. *Cell* **140**, 268–279 (2010).
- Blumenthal, A. *et al.* The Wingless homolog WNT5A and its receptor Frizzled-5 regulate inflammatory responses of human mononuclear cells induced by microbial stimulation. *Blood* **108**, 965–973 (2006).
- Stockmann, C. *et al.* Deletion of vascular endothelial growth factor in myeloid cells accelerates tumorigenesis. *Nature* **456**, 814–818 (2008).
- Lin, S. L. *et al.* Macrophage Wnt7b is critical for kidney repair and regeneration. *Proc. Natl Acad. Sci. USA* **107**, 4194–4199 (2010).
- Tauber, A. I. Metchnikoff and the phagocytosis theory. *Nature Rev. Mol. Cell Biol.* **4**, 897–901 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Speeg for technical assistance and A. P. McMahon for the *Wnt11* mice. This work was supported by the NIH (J.A.S., M.W.K., J.W.P., J.D.M., T.Y., B.O.W., R.A.L.) by the HHMI (J.D.M.) and Cancer Research UK (H.G.).

Author Contributions R.A.L. provided project leadership and wrote the manuscript with J.A.S., J.A.S., I.L., S.R., H.G., and R.A.L. designed the experiments. J.A.S., I.L., S.R., G.M., A.C.C., A.R.B., J.F., and R.A. performed the experiments. S.R., J.W.P., T.Y., N.F. and B.O.W. developed critical reagents. Experimental supervision and helpful discussions were provided by M.W.K., J.D.M., S.R., J.W.P., and H.G.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.A.L. (richard.lang@cchmc.org).

METHODS

Quantification of retinal vasculature and isolation of RMCs. Retinas were prepared and imaged as reported¹ except that Alexa Fluor 488 isolectin GS-IB₄ (Invitrogen) was used to label retinal vessels and RMCs. Other antibodies included anti-endomucin (Santa Cruz, V.7C7), anti-GFP (Abcam), anti-Iba1 (Wako), anti-CD204 (AbD Serotec), and anti-Flt1 (R&D). For quantification, we used 200× magnification images located at the retinal periphery between artery and vein. For each genotype, at least three fields were analysed from at least four animals from a minimum of two litters. Control and experimental animals were littermates. This minimized the effect of strain background in producing variation in vascular density^{31–34}. For flow sorting, retinas were digested and sorted as reported¹⁶ except that 0.5 mg ml^{−1} DNase II (Sigma Aldrich) was used. Furthermore, myeloid cells were pre-enriched using CD11b beads (Miltenyi Biotec). Cells were then incubated with anti-CD16/32 (clone 24.G2) for 30 min and labelled with monoclonal antibodies to phycoerythrin-Cy7 conjugated anti-CD11b (clone M1/70), allophycocyanin-Cy7 or peridinin chlorophyll A protein-Cy5.5 conjugated anti-F4/80 (clone BM8), Alexa-647 conjugated anti-CD204 (clone 2F8), and 7-aminoactinomycin D. Cells were sorted with a FACSAria II running DiVa software.

RNA isolation and RT-PCR. RNA was isolated using RNeasy (Qiagen). Quantitative PCR was performed with QuantiTect SYBR green (Qiagen). Primers are listed in Supplementary Table 1.

In vitro analysis. MEFs were isolated as described³⁵ from *Wls^{fl/fl}* mice³⁶ and transfected with combinations of 2.5 µg Wnt5a, Thy1.1 and cre plasmids using TransIT-2020 (Mirus). MEFs were sorted using magnetic beads for Thy1.1, replated and supernatant was collected after 24 h. RAW264.7 cells (ATCC) were grown in DMEM (10% FBS) and treated with recombinant Wnt3a (10 ng ml^{−1}, R&D) or Wnt5a (500 ng ml^{−1}, R&D). ELISA was performed using the sVEGFR1 Quantikine kit (R&D). For calcium imaging, RAW264.7 cells were loaded in

Riger's solution with 5 µM Fura-2 AM and imaged at 510 nm at 1 Hz after excitation at 340 nm and 380 nm. Data was acquired using EasyRatioPro. Ca²⁺ traces are 50-cell averages from two experiments.

Statistics. All statistical tests used are stated in the figure legends. In analysing quantitative PCR data, the *P* values refer to a comparison of the $\Delta\Delta C_t$ values.

Animals. Breeding and genotyping of *Wls^{fl}* (ref. 36), *cfms-icre* (ref. 18), *Z/EG* (ref. 37), *Wnt5a^{+/-}* (ref. 19), *Wnt11^{+/-}* (ref. 20) and *Flt^{fl}* (ref. 26) was performed as previously described. *Lrp5^{fl}* will be described in detail in a forthcoming publication. The allele is a conventional design where exon 2 of the *Lrp5* gene is flanked by LoxP sites. It has been confirmed that deletion between the LoxP sites produces a loss-of-function. All animal experimentation was carried out using protocols approved by the Institutional Animal Care and Use Committee.

31. Rohan, R. M., Fernandez, A., Udagawa, T., Yuan, J. & D'Amato, R. J. Genetic heterogeneity of angiogenesis in mice. *FASEB J.* **14**, 871–876 (2000).
32. Gao, G. *et al.* Difference in ischemic regulation of vascular endothelial growth factor and pigment epithelium-derived factor in Brown Norway and Sprague Dawley rats contributing to different susceptibilities to retinal neovascularization. *Diabetes* **51**, 1218–1225 (2002).
33. Chan, C. K. *et al.* Mouse strain-dependent heterogeneity of resting limbal vasculature. *Invest. Ophthalmol. Vis. Sci.* **45**, 441–447 (2004).
34. Chan, C. K. *et al.* Differential expression of pro- and antiangiogenic factors in mouse strain-dependent hypoxia-induced retinal neovascularization. *Lab. Invest.* **85**, 721–733 (2005).
35. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. *Manipulating the mouse embryo: a laboratory manual*. 3rd edn, 371–373 (Cold Spring Harbor Laboratory Press, 2003).
36. Carpenter, A. C., Rao, S., Wells, J. M., Campbell, K. & Lang, R. A. Generation of mice with a conditional null allele for *Wntless*. *Genesis* **48**, 554–558 (2010).
37. Novak, A., Guo, C., Yang, W., Nagy, A. & Lobe, C. G. *Z/EG*, a double reporter mouse line that expresses enhanced green fluorescent protein upon Cre-mediated excision. *Genesis* **28**, 147–155 (2000).

A current filamentation mechanism for breaking magnetic field lines during reconnection

H. Che¹, J. F. Drake² & M. Swisdak³

During magnetic reconnection, the field lines must break and reconnect to release the energy that drives solar and stellar flares^{1,2} and other explosive events in space³ and in the laboratory⁴. Exactly how this happens has been unclear, because dissipation is needed to break magnetic field lines and classical collisions are typically weak. Ion–electron drag arising from turbulence⁵, dubbed ‘anomalous resistivity’, and thermal momentum transport⁶ are two mechanisms that have been widely invoked. Measurements of enhanced turbulence near reconnection sites in space^{7,8} and in the laboratory^{9,10} support the anomalous resistivity idea but there has been no demonstration from measurements that this turbulence produces the necessary enhanced drag¹¹. Here we report computer simulations that show that neither of the two previously favoured mechanisms controls how magnetic field lines reconnect in the plasmas of greatest interest, those in which the magnetic field dominates the energy budget. Rather, we find that when the current layers that form during magnetic reconnection become too intense, they disintegrate and spread into a complex web of filaments that causes the rate of reconnection to increase abruptly. This filamentary web can be explored in the laboratory or in space with satellites that can measure the resulting electromagnetic turbulence.

Particle-in-cell simulations reveal that the rate of reconnection in a three-dimensional system jumps sharply above the rate measured in two dimensions (Fig. 1a). The jump is a consequence of the development of turbulence in three dimensions. To understand why this happens, it is necessary to explore how magnetic fields break and reconnect to form the topological x -line that characterizes reconnection (Fig. 2a). Magnetic reconnection produces large electric fields that are parallel to the local magnetic field in the vicinity of the x -line (Fig. 2b). In the absence of a dissipative process, such a parallel electric field would produce an infinite current and reconnection would cease. The mechanisms that limit the electron response to parallel electric fields therefore break field lines. These mechanisms can be understood from the electron momentum equation in the direction perpendicular to the plane of reconnection (the z direction)¹²:

$$\frac{\partial p_{ez}}{\partial t} + \nabla \cdot \mathbf{p}_e \mathbf{v}_{ez} = -en_e E_z + \frac{1}{c}(\mathbf{j}_e \times \mathbf{B})_z - \nabla \cdot \mathbf{P}_{ez} \quad (1)$$

where \mathbf{p}_e is the momentum density, \mathbf{j}_e is the electron current density and $\mathbf{P}_{ez} = P_e \hat{z}$ with P_e the pressure tensor. In a three-dimensional system where the intense current layers produced during reconnection (Fig. 2c) drive short-scale turbulence with wavevectors along z , the impact of this turbulence can be quantified by averaging equation (1) over z to obtain a generalized Ohm’s law:

$$\begin{aligned} \langle E_z \rangle = & -\frac{1}{c}(\langle \mathbf{v}_{e\perp} \rangle \times \langle \mathbf{B}_{\perp} \rangle)_z - \frac{\nabla_{\perp} \cdot \langle \mathbf{P}_{ez} \rangle}{\langle n_e \rangle e} \\ & - \frac{m_e}{e} \left(\frac{\partial \langle v_{ez} \rangle}{\partial t} + \langle \mathbf{v}_{e\perp} \rangle \cdot \nabla_{\perp} \langle v_{ez} \rangle \right) + D_{ez} + \nabla_{\perp} \cdot \mathbf{T}_{ez} \end{aligned}$$

where $\langle f \rangle$ denotes an average of f over z with $\delta f = f - \langle f \rangle$, and the subscript \perp denotes the x - y plane.

$$D_{ez} = \frac{-1}{\langle n_e \rangle} \langle \delta n_e \delta E_z \rangle$$

and

$$\mathbf{T}_{ez} = -\frac{1}{\langle n_e \rangle e} \langle \delta \mathbf{p}_{e\perp} \left(\delta v_{ez} - \frac{e}{m_e c} \delta A_z \right) \rangle$$

are the turbulent drag and momentum transport, respectively, and δA_z is the vector potential for $\delta \mathbf{B}_{\perp}$. This representation for \mathbf{T}_{ez} is valid only in the strong guide field limit (see the Supplementary Information). In a two-dimensional system D_{ez} and \mathbf{T}_{ez} are both zero and $\mathbf{v}_{e\perp}$ is typically zero at the x -line. In steady state the inertia term is also zero, so only the thermal momentum transport arising from the xz and yz components of \mathbf{P}_{ez} can balance the reconnection electric field. This was confirmed in earlier two-dimensional simulations when the ambient guide magnetic field B_z is small or absent⁶. However, in systems in which the magnetic energy exceeds the local electron plasma pressure so that $\beta_e = 8\pi n_e T_e / B^2$ is small, such as in solar and stellar coronae and some regions of the Earth’s magnetosphere, we show that thermal momentum transport is weak. Therefore, in such systems other mechanisms for balancing the reconnection electric field need to be explored.

In a three-dimensional system the strong currents driven by the reconnection electric field around the x -line could potentially drive turbulence, scattering electrons to produce an electron–ion drag or anomalous resistivity to balance the reconnection electric field^{13,12–14}. This turbulence is typically weak or absent in two dimensions because the growth rates of instabilities peak when wavevectors have components along the direction of the current, which is strongest along z . The drag D_{ez} in equation (2) measures the effect of this turbulence but neither simulations nor observations have established the viability of the mechanism. The narrow current layers that result from reconnection can also break up through self-driven turbulence. The resulting turbulent transport of electron momentum away from the x -line as described by \mathbf{T}_{ez} could balance the reconnection electric field and act to break field lines^{15,16}. We explore this idea below.

At larger spatial scales ambient turbulence, which is common in astrophysical plasmas, might enable fast reconnection by breaking up the large-scale current layers that normally characterize magnetohydrodynamic reconnection^{17,18}. In such models, however, the question of what provides the dissipation required to break field lines is not addressed.

The development of turbulence during simulations of three-dimensional reconnection and its consequences are presented in Figs 1–4. A comparison of the time-development of the rate of reconnection between two and three dimensions reveals that for a system that initially has low β_e the rate of reconnection rises over time, with the two simulations matching until the rate in the three-dimensional case abruptly increases (Fig. 1a). Coincident with the rate increase in three dimensions is an increase in the turbulent-driven drag and momentum transport (Fig. 1b). The reconnection electric field E_z drives a strong current around the x -line that strengthens and narrows over time (Fig. 1a). At late time the current layer in the three-dimensional case

¹Center For Integrated Plasma Studies, Department of Physics, University of Colorado, Boulder, Colorado 80309, USA. ²Department of Physics and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA. ³Department of Physics, University of Maryland, College Park, Maryland 20742, USA.

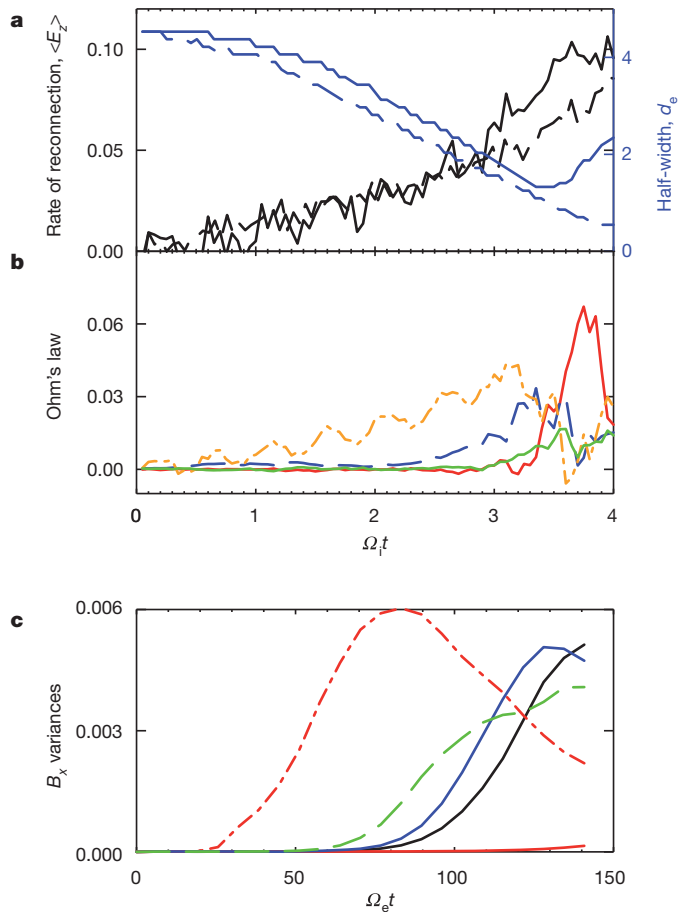


Figure 1 | The time evolution of reconnection and the development of turbulence. Particle-in-cell simulations using the p3d code²⁷ are performed in doubly periodic two-dimensional and three-dimensional geometries starting with two force-free current sheets. The reconnection magnetic field is $B_x/B_0 = \tanh[(y - L_y/4)/w_0] - \tanh[(y - 3L_y/4)/w_0] - 1$, where the current layer width $w_0 = 0.5d_i$ and the box size is $L_x \times L_y \times L_z = 4d_i \times 2d_i \times 8d_i$. The electron and ion temperatures, $T_e/m_e c_A^2 = T_i/m_i c_A^2 = 0.04$, and density n_0 are initially uniform. The initial out-of-plane 'guide' field B_z/B_0 is 5.0 outside the current layers and increases within the current layers so that the total magnetic field B is a constant. The cyclotron time is $\Omega_i^{-1} = m_i c / e B_0$, the Alfvén speed is $c_A = B_0 / (4\pi m_i n_0)^{1/2}$ and the ion inertial length is $d_i = c_A / \Omega_i$. The electron mass is $0.01m_i$ and the velocity of light $c = 20c_A$. The three-dimensional spatial grid consists of $512 \times 256 \times 1,024$ cells with 20 particles per cell, whereas in two dimensions it consists of $2,048 \times 1,024$ with 100 particles per cell. Reconnection is initiated with a large-scale magnetic perturbation. In **a** the rates of reconnection $\langle E_z \rangle$ (normalized to $E_0 = c_A B_0 / c$) and the half-widths of the z -averaged electron current layer at the x -line (see Fig. 2a) in three dimensions (solid lines) and two dimensions (dashed lines) are shown. In **b** the dominant components of Ohm's law (normalized to E_0), the electron inertia $-(m_e/e)\partial \langle v_{ez} \rangle / \partial t$ (orange line), the drag D_z (blue line), the y -directed turbulent momentum transport $\partial T_{yz} / \partial y$ (red line) and the thermal momentum transport $-\nabla_\perp \cdot \langle \mathbf{P}_{ez} \rangle / \langle n_e \rangle e$ (green line) are shown. In three-dimensional simulations with four times the initial electron and ion temperatures (not shown), no turbulence develops and thermal momentum transport balances $\langle E_z \rangle$. Finally, in **c** the magnetic field variances $\langle \delta B_x^2 \rangle / B_0^2$ are plotted versus time from a series of three-dimensional simulations of narrow current layers (no reconnection) with various guide fields B_z/B_0 (5.0 in black, 2.5 in green, 1.0 in blue and 0.5 in red) and stationary ions are shown. The simulation domains are $L_x \times L_y \times L_z = 10d_e \times 10d_e \times 80d_e$ and all but the dot-dashed red line have current layer widths of d_e . All but the guide field of 0.5 break up, but reducing the layer width in this case from $1.0d_e$ to $0.5d_e$ (dot-dashed red line) again leads to break-up. Current layers with Harris equilibria also break up (see the Supplementary Information).

abruptly broadens as the nearly laminar current layer undergoes a transition to a filamented current layer (Fig. 3a and b). The sharp jump in the strength of the magnetic field perturbations (Fig. 3c and d) and

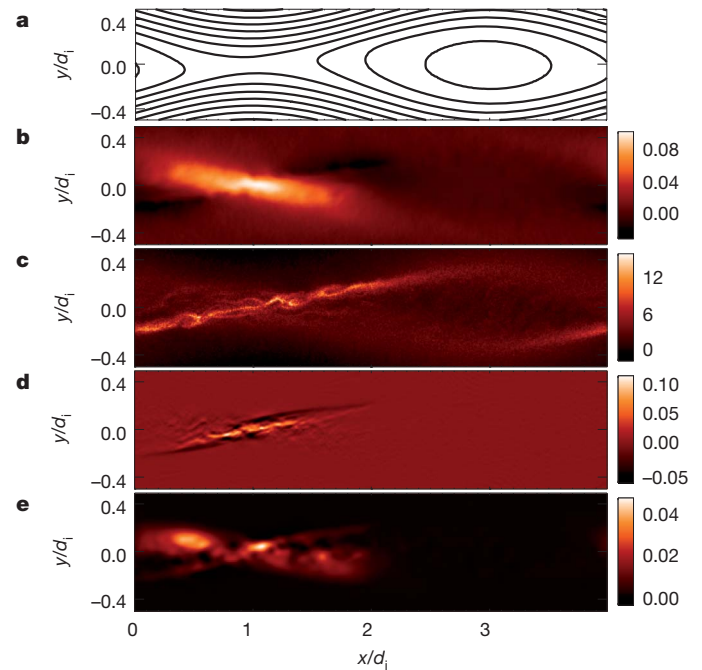


Figure 2 | The geometry of magnetic reconnection at late time. Shown is the x - y plane at $t = 3.75 \Omega_i^{-1}$ from the three-dimensional simulation of Fig. 1. In **a** the reconnected magnetic field lines (averaged over z) are shown. In **b** the field line motion towards and away from the x -line induces an electric field that produces the parallel (to \mathbf{B}) electric field $\langle E_\parallel \rangle$ that drives the intense electron current J_{ez} around the x -line in **c**. The irregular structure of the current layer indicates that it is turbulent (see also Fig. 3). The filamentation of the current layer transports electron momentum p_{ez} away from the centre of the current layer. The rate of momentum transport $\nabla \cdot \mathbf{T}_{ez}$ is shown in **d**. The positive peak in $\nabla \cdot \mathbf{T}_{ez}$ with negative values upstream demonstrates that momentum is being transported away from the x -line. The turbulence also produces a net electron-ion drag D_z , shown in **e**. We note that the spatial structure of the drag is very different from the turbulent momentum transport. It does not overlap with the region of high current density and therefore cannot be represented as an anomalous resistivity. The colour scales measure the variables normalized to E_0 (**b**, **d** and **e**) and $n_0 e c_A$ (**c**).

the wavelength of the electric field perturbations (Fig. 3e and f) indicate that this filamentation instability is distinct from instabilities explored earlier in observations^{7,8} and modelling^{19–23}.

To explore the role of this turbulence in reconnection and determine whether it facilitates the breaking of field lines, we evaluate Ohm's law in equation (2) around the x -line (Fig. 4). In two dimensions the reconnection electric field is balanced by the electron inertia even at late time. In three dimensions the mechanisms controlling how the magnetic field lines break evolve over time: early in time it is the electron inertia, at intermediate times the drag reaches parity with the inertia, and at late time the turbulent momentum transport dominates. The broadening of the current layer (Fig. 1a) causes the electron streaming velocity, the inertia term in Ohm's law and the turbulent drag to decrease sharply at the x -line (Fig. 1b). By the end of the simulation the momentum transport has also decreased because of the broadening of the current layer. In larger simulation domains, where more magnetic flux is available to reconnect, the process of current layer thinning and broadening goes through several cycles before coming to an approximate balance (see the movie in the Supplementary Information).

What drives the filamentation of the current shown in Fig. 3? The time-dependence of the transverse electric fields reveals that it is a right-hand circularly polarized electromagnetic wave and hence is part of the whistler/electron-cyclotron branch. The spatial correlation between the current density and transport in Fig. 2 suggests that it is driven by the current density gradient and not by the relative streaming velocity of electrons and ions. Three-dimensional particle-in-cell

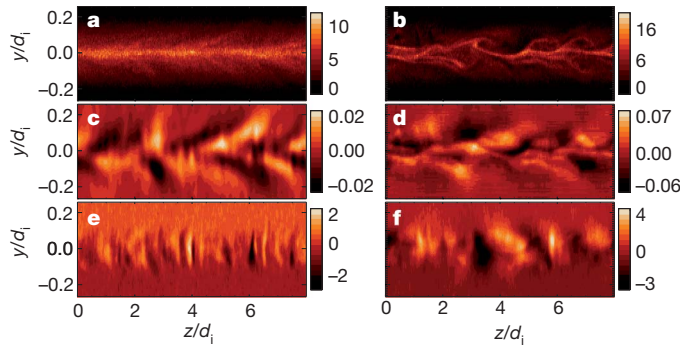


Figure 3 | The filamentary structure of the electron current layer. In the y - z plane in a cut through the x -line at $t = 3.0\Omega_{ci}^{-1}$ (a, c and e) and $t = 3.75\Omega_{ci}^{-1}$ (b, d and f) are the electron current j_{ez}/n_0ec_A (a and b), the magnetic field perturbations $\delta B_x/B_0$ (c and d) and the electric field perturbations $\delta E_x/E_0$ (e and f). The filamentation instability begins sharply at $t = 3.5\Omega_{ci}^{-1}$. This instability did not appear in earlier simulations^{13,14,28} because the computational domains along the z axis were too short to capture its long wavelength parallel to \mathbf{B} (in ref. 28 the computational domain along the z axis calculated from the upstream density was only $2.86d_i$). The measured phase speed of the filaments is close to the electron drift speed, suggesting that the coupling to the ions is weak. The colour scales measure the normalized variables.

simulations of narrow current layers reveal that the filamentation instability is insensitive to the ion mass and remains robust even for stationary ions (see Supplementary Information). A simple fluid description of electromagnetic waves in this regime is given by the electron magnetohydrodynamic equations²⁴ and the possibility that gradients in the electron current could drive instability has been previously discussed^{16,25}. Linearization of the electron magnetohydrodynamic

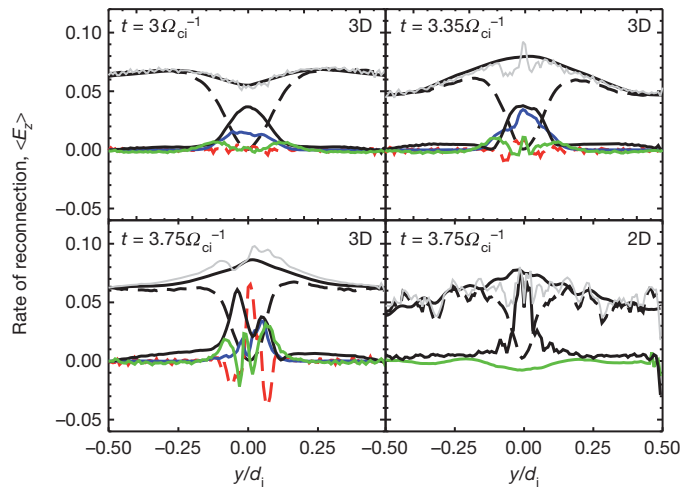


Figure 4 | Breaking magnetic field lines: the dominant components of Ohm's law. At $t = 3.0\Omega_{ci}^{-1}$, $t = 3.35\Omega_{ci}^{-1}$ and $t = 3.75\Omega_{ci}^{-1}$ from the three-dimensional (3D) simulation and $t = 3.75\Omega_{ci}^{-1}$ from the two-dimensional (2D) simulation, the dominant contributions to Ohm's law are plotted in cuts along the inflow direction y through the x -line. The lines shown are $\langle E_z \rangle$ (upper solid black), $-v_{ey} \langle B_x \rangle / c$ (dashed black), electron inertia (lower solid black), the thermal momentum transport $-(1/\langle n_e \rangle e) \partial \langle P_{eyz} \rangle / \partial y$ (solid green), the drag D_z (blue) and turbulent transport $\partial T_{yz} / \partial y$ (red). The solid grey line shows the sum of all of the contributions to Ohm's law, which should lie on top of $\langle E_z \rangle$. In two dimensions the electron inertia continues to balance the reconnection electric field at late time, a solution that is not consistent with steady reconnection. In three dimensions the turbulent drag and inertia dominate at $t = 3.0\Omega_{ci}^{-1}$. By $t = 3.35\Omega_{ci}^{-1}$ the turbulent drag and the rate of reconnection both sharply increase, suggesting a causal relation. By $t = 3.75\Omega_{ci}^{-1}$, the current layer is becoming filamentary and turbulent momentum transport completely dominates the force balance at the x -line. Momentum is transported upstream away from the x -line, producing a positive $\partial T_{yz} / \partial y$ at the x -line and negative values upstream corresponding to a momentum transfer, not a momentum sink. At the x -line the drag drops sharply as the dynamics of the filaments dominates.

equations in the presence of a local current density gradient $J_{ez}' = dJ_{ez}/dy$ yields an instability with a peak growth rate of $\gamma \approx J_{ez}' / (2ne)$ (see Supplementary Information). Of course, a kinetic treatment will be required to understand this instability fully.

What is the range of parameters (for example, the guide field) over which the instability is important? A series of three-dimensional particle-in-cell simulations of narrow current layers reveal a surprising result: the guide field has a destabilizing rather than a stabilizing influence (Fig. 1c). The instability requires a finite guide field to develop, is strongest for a guide field that is around twice the reconnecting magnetic field, but remains robust down to a guide field of 0.5 of the reconnecting field for current layer widths of $0.5d_e$, where d_e is the electron skin depth (see also the Supplementary Information). During reconnection in a two-dimensional system with a guide field, the width of the current layer decreases to the electron Larmor radius $\rho_e = \beta_e^{0.5} d_e$ (refs 6 and 26). Thus, in the Earth's magnetosphere, reconnection-driven current layers should filament for values of β_e below 0.25 and guide fields around 0.5.

Satellites such as Cluster and THEMIS should be able to measure the predicted asymmetry in the distribution of turbulence between the two separatrices (Fig. 2c, d). Detailed measurements of d_e -scale current layers must await the results of the Magnetospheric Multiscale Mission. Filamentation of current layers in the Versatile Toroidal Facility reconnection experiment at the Massachusetts Institute of Technology has already been observed (J. Egedal, personal communication).

Received 22 June 2010; accepted 1 April 2011.

Published online 1 June 2011.

1. Tsuneta, S. Heating and acceleration processes in hot thermal and impulsive solar flares. *Astrophys. J.* **290**, 353–358 (1985).
2. Priest, E. R. & Forbes, T. G. *Magnetic Reconnection: MHD Theory and Applications* (Cambridge University Press, 2000).
3. Baker, D. N., Pulkkinen, T. I., Angelopoulos, V., Baumjohann, W. & McPherron, R. L. Neutral line model of substorms: past results and present view. *J. Geophys. Res.* **101**, 12975–13010 (1996).
4. Yamada, M. *et al.* Investigation of magnetic reconnection during a sawtooth crash in a high-temperature tokamak plasma. *Phys. Plasmas* **1**, 3269–3276 (1994).
5. Galeev, A. A. & Sagdeev, R. Z. in *Basic Plasma Physics* (eds Galeev, A. A. & Sudan, R. N.) Vol. 1, 677–731 (North Holland Publishing Company, 1983).
6. Hesse, M., Kuznetsova, M. & Hoshino, M. The structure of the dissipation region for component reconnection: particle simulations. *Geophys. Res. Lett.* **29**, 1563, doi:10.1029/2001GL014714 (2002).
7. Matsumoto, H., Deng, X. H., Kojima, H. & Anderson, R. R. Observation of electrostatic solitary waves associated with reconnection on the dayside magnetopause boundary. *Geophys. Res. Lett.* **30**, 1326, doi:10.1029/2002GL016319 (2003).
8. Cattell, C. *et al.* Cluster observations of electron holes in association with magnetotail reconnection and comparison to simulations. *J. Geophys. Res.* **110**, A01211 (2005).
9. Ji, H. *et al.* Electromagnetic fluctuations during fast reconnection in a laboratory plasma. *Phys. Rev. Lett.* **92**, 115001 (2004).
10. Fox, W. *et al.* Laboratory observation of electron phase-space holes during magnetic reconnection. *Phys. Rev. Lett.* **101**, 255003 (2008).
11. Eastwood, J., Phan, T. D., Bale, S. D. & Tjulin, A. Observations of turbulence generated by magnetic reconnection. *Phys. Rev. Lett.* **102**, 035001 (2009).
12. Vasyliunas, V. M. Theoretical models of magnetic field line merging. 1. *Rev. Geophys. Space Phys.* **13**, 303–336 (1975).
13. Drake, J. F. *et al.* Formation of electron holes and particle energization during magnetic reconnection. *Science* **299**, 873–877 (2003).
14. Che, H., Drake, J. F., Swisdak, M. & Yoon, P. H. Electron holes and heating in the reconnection dissipation region. *Geophys. Res. Lett.* **37**, L11105, doi:10.1029/2010GL043608 (2010).
15. Kaw, P. K., Valeo, E. J. & Rutherford, P. H. Tearing modes in a plasma with magnetic braiding. *Phys. Rev. Lett.* **43**, 1398–1401 (1979).
16. Drake, J. F., Kleva, R. G. & Mandt, M. E. Structure of thin current layers: implications for magnetic reconnection. *Phys. Rev. Lett.* **73**, 1251–1254 (1994).
17. Lazarian, A. & Vishniac, E. Reconnection in a weakly stochastic field. *Astrophys. J.* **517**, 700, doi:10.1086/307233 (1999).
18. Kowal, G., Lazarian, A., Vishniac, E. & Otmianowska-Mazur, K. Reconnection in a weakly stochastic magnetic field. *Astrophys. J.* **700**, 63–85 (2009).
19. Openheim, M. M., Vetoulis, G., Newman, D. L. & Goldman, M. V. Evolution of electron phase-space holes in 3-D. *Geophys. Res. Lett.* **28**, 1891–1894 (2001).
20. Omura, Y., Matsumoto, H., Miyake, T. & Kojima, H. Electron beam instabilities as a generation mechanism of electrostatic solitary waves in the magnetotail. *J. Geophys. Res.* **101**, 2685–2697 (1996).
21. McMillan, B. F. & Cairns, I. H. Lower hybrid turbulence driven by parallel currents and associated electron energization. *Phys. Plasmas* **13**, 052104, doi:10.1063/1.2198212 (2006).

22. Goldman, M. V., Newman, D. L. & Pritchett, P. Vlasov simulations of electron holes driven by particle distribution from PIC reconnection simulations with a guide field. *Geophys. Res. Lett.* **35**, L22109, doi:10.1029/2008GL035608 (2008).
23. Che, H., Drake, J. F., Swisdak, M. & Yoon, P. H. Nonlinear development of streaming instabilities in strongly magnetized plasma. *Phys. Rev. Lett.* **102**, 145004 (2009).
24. Kingsep, A. S., Chukbar, K. V. & Yan'kov, Y. Y. in *Reviews of Plasma Physics* (ed. Kadomtsev, B. B.) Vol. 16, 243–288 (Consultants Bureau, 1990).
25. Ferraro, N. M. & Rogers, B. N. Turbulence in low- β reconnection. *Phys. Plasmas* **11**, 4382–4389 (2004).
26. Ricci, P., Brackbill, J. U., Daughton, W. & Lapenta, G. Collisionless magnetic reconnection in the presence of a guide field. *Phys. Plasmas* **11**, 4102–4114 (2004).
27. Zeiler, A. *et al.* Three-dimensional particle simulations of collisionless magnetic reconnection. *J. Geophys. Res.* **107**, 1230, doi:10.1029/2001JA000287 (2002).
28. Pritchett, P. & Coroniti, F. V. Three-dimensional collisionless magnetic reconnection in the presence of a guide field. *J. Geophys. Res.* **109**, A01220, doi:10.1029/2003JA009999 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work has been supported by the NSF/DOE programme in plasma science and by NASA through the Supporting Research and Technology Program and the Magnetospheric Multiscale Mission Science Team. Computations were carried out in part at the National Energy Research Scientific Computing Center.

Author Contributions All of the authors made significant contributions to this work. H.C. carried out the particle simulations of reconnection. M.S. carried out simulations of isolated electron current layers. H.C., J.F.D. and M.S. analysed the data from the simulations. All authors discussed the results and commented on the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.C. (haihong.che@colorado.edu).

Latent TGF- β binding protein 3 identifies a second heart field in zebrafish

Yong Zhou^{1,2}, Timothy J. Cashman^{1,2}, Kathleen R. Nevis^{1,2}, Pablo Obregon^{1,2}, Sara A. Carney³, Yan Liu^{2,4}, Aihua Gu^{1,2,5}, Christian Mosimann^{2,6}, Samuel Sondalle^{1,2,7}, Richard E. Peterson³, Warren Heideman³, Caroline E. Burns^{1,2,7} & C. Geoffrey Burns^{1,2}

The four-chambered mammalian heart develops from two fields of cardiac progenitor cells distinguished by their spatiotemporal patterns of differentiation and contributions to the definitive heart^{1–3}. The first heart field differentiates earlier in lateral plate mesoderm, generates the linear heart tube and ultimately gives rise to the left ventricle. The second heart field (SHF) differentiates later in pharyngeal mesoderm, elongates the heart tube, and gives rise to the outflow tract and much of the right ventricle. Because hearts in lower vertebrates contain a rudimentary outflow tract but not a right ventricle⁴, the existence and function of SHF-like cells in these species has remained a topic of speculation^{4–10}. Here we provide direct evidence from Cre/Lox-mediated lineage tracing and loss-of-function studies in zebrafish, a lower vertebrate with a single ventricle, that *latent TGF- β binding protein 3* (*ltbp3*) transcripts mark a field of cardiac progenitor cells with defining characteristics of the anterior SHF in mammals. Specifically, *ltbp3*⁺ cells differentiate in pharyngeal mesoderm after formation of the heart tube, elongate the heart tube at the outflow pole, and give rise to three cardiovascular lineages in the outflow tract and myocardium in the distal ventricle. In addition to expressing *Ltbp3*, a protein that regulates the bioavailability of TGF- β ligands¹¹, zebrafish SHF cells co-express *nkx2.5*, an evolutionarily conserved marker of cardiac progenitor cells in both fields⁴. Embryos devoid of *ltbp3* lack the same cardiac structures derived from *ltbp3*⁺ cells

due to compromised progenitor proliferation. Furthermore, small-molecule inhibition of TGF- β signalling phenocopies the *ltbp3*-morphant phenotype whereas expression of a constitutively active TGF- β type I receptor rescues it. Taken together, our findings uncover a requirement for *ltbp3*-TGF- β signalling during zebrafish SHF development, a process that serves to enlarge the single ventricular chamber in this species.

Using *in situ* hybridization, we discovered that zebrafish *ltbp3* transcripts are expressed in cells at the outflow pole of the linear heart tube (Fig. 1a, b and Supplementary Figs 1 and 2). Double marker analyses confirmed that a majority of *ltbp3*⁺ cells are non-overlapping with differentiated myocardium (Fig. 1c–e and Supplementary Fig. 2). Additionally, *ltbp3*⁺ cells are neither endothelial nor derived from the neural crest because *ltbp3* expression remains robust in embryos lacking both cell types (Supplementary Fig. 3). Because the anatomical relationship of *ltbp3*⁺ cells to the heart tube is reminiscent of the anterior segment of the SHF in mice^{3,12}, *ltbp3*⁺ cells were evaluated for co-expression of *nkx2.5*, an evolutionarily conserved marker of cardiac progenitor cells (CPCs) in both heart fields⁴. As reported previously, cells in the linear heart tube were positive for *nkx2.5* (ref. 13). Surprisingly, *nkx2.5* expression also overlapped with *ltbp3* transcripts (Fig. 1f–h), demonstrating that an extra-cardiac population of *ltbp3*⁺, *nkx2.5*⁺ cells resides at the outflow pole of the zebrafish heart tube. This population was also readily identified in double transgenic

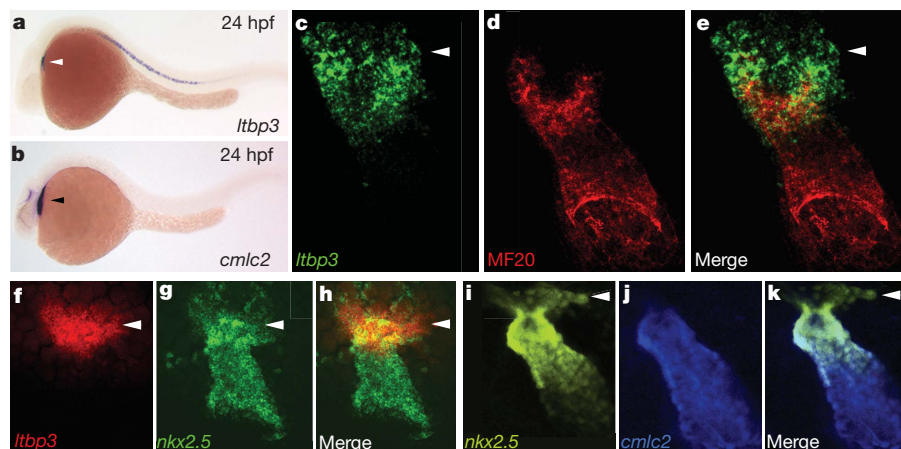


Figure 1 | *ltbp3* and *nkx2.5* transcripts mark extra-cardiac cells contiguous to the outflow pole of the zebrafish heart tube. **a, b**, *ltbp3*⁺ cells (white arrowhead) visualized by *in situ* hybridization at 24 h post-fertilization (hpf) reside dorsal to the *cmhc2*⁺ heart tube (black arrowhead; *n* > 15 embryos per group). **c–e**, Heart tube region in 24 hpf embryo co-stained with *ltbp3*⁺ riboprobe (green; arrowhead) and a muscle-specific antibody (MF20, red;

n = 3/3) that recognizes cardiomyocytes. **f–h**, Heart tube region in 24 hpf embryo co-stained with *ltbp3* and *nkx2.5* riboprobes highlighting *ltbp3*⁺, *nkx2.5*⁺ cells (arrowheads) at the outflow pole of the heart tube (*n* = 9/9). **i–k**, Heart tube region in 26 hpf *Tg(nkx2.5::ZsYellow)*; *Tg(cmhc2::CSY)* embryo highlighting non-myocardial *nkx2.5*⁺ cells (arrowheads). Original magnifications: **a, b** 4 \times ; **c–k**, 40 \times .

¹Cardiovascular Research Center, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA. ²Harvard Medical School, Boston, Massachusetts 02115, USA. ³Pharmaceutical Sciences Division, School of Pharmacy, University of Wisconsin, Madison, Wisconsin 53705, USA. ⁴Nephrology Division, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA. ⁵School of Public Health, Nanjing Medical University, Nanjing, 210029, China. ⁶Stem Cell Program and Division of Hematology/Oncology, Children's Hospital Boston, Massachusetts 02115, USA. ⁷Harvard Stem Cell Institute, Cambridge, Massachusetts 02138, USA.

Tg(nkx2.5::ZsYellow); *Tg(cmlc2::CSY)* embryos expressing ZsYellow and AmCyan proteins from *nkx2.5* and myocardial (*cmlc2*) promoters respectively (Fig. 1i–k). Lastly, these cells co-express zebrafish *tgfb3* transcripts (Supplementary Fig. 4) consistent with the demonstrated function of LTBP proteins as regulators of TGF- β signalling¹¹.

To determine if *Ltbp3* is required for zebrafish cardiogenesis, we evaluated embryos injected with anti-sense *ltbp3* morpholinos (Supplementary Fig. 5) for chamber-specific defects in cardiomyocyte and endocardial cell number. Knocking down *Ltbp3* halved the number of ventricular cardiomyocytes and endocardial cells, whereas atrial cell numbers were unaffected (Fig. 2a–f). The ventricular deficit was evident earlier in development, soon after formation of the heart tube, as shortening of the ventricular segment accompanied by a defect in cardiac looping (Fig. 2g–i). *ltbp3* morphants also lacked *Eln2*⁺ outflow tract (OFT) smooth muscle precursor cells (Fig. 2j–m)^{7,14} homologous to SHF-derived smooth muscle cells surrounding the base of the aorticopulmonary trunk in higher vertebrates¹⁰. Lastly, *ltbp3* morphants failed to form the ventral aorta (data not shown) and a majority of pharyngeal arch arteries (Supplementary Fig. 6). Taken together, these data demonstrate that knocking down *ltbp3* causes multi-lineage cardiovascular defects in the pharyngeal arches, OFT and ventricle.

In mammals, the anterior segment of the SHF gives rise to the embryonic OFT and the right ventricular half of the common embryonic ventricle before septation^{1–3}. Despite the fact that zebrafish embryos never septate their single ventricles, the *ltbp3* morphant

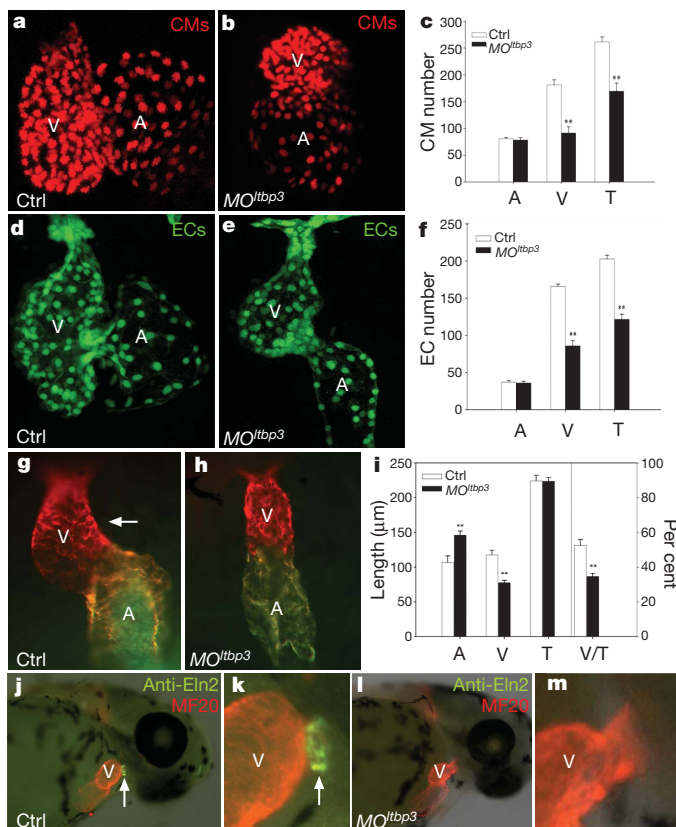


Figure 2 | Knocking down *Ltbp3* causes multi-lineage cardiovascular defects in the ventricle and OFT. a–f, Cardiomyocyte (CM) and endocardial (EC) cell numbers in control and *ltbp3* morphant (*MO^{ltbp3}*) atria (A) and ventricles (V) at 72 hpf in *cmlc2::dsRed2-nuc* and *fli1::nEGFP* embryos ($n = 6$ per group, $T = \text{total}$). g–i, Atrial (A) and ventricular (V) segment lengths in control and *ltbp3* morphant embryos at 36 hpf ($n = 9$ per group; white arrow highlights rightward looping of the ventricle in control embryos). Error bars in all graphs represent 1 s.d., ** $P < 0.01$. j–m, 60 hpf control and *MO^{ltbp3}* embryos stained with anti-*Eln2* antibodies ($n = 24/24$ for wild type; $28/31$ for *MO^{ltbp3}*) that recognize OFT smooth muscle cell precursor cells (white arrow). Original magnifications: a, b, d, e, 40 \times ; g, h, 10 \times ; j–m, 4 \times .

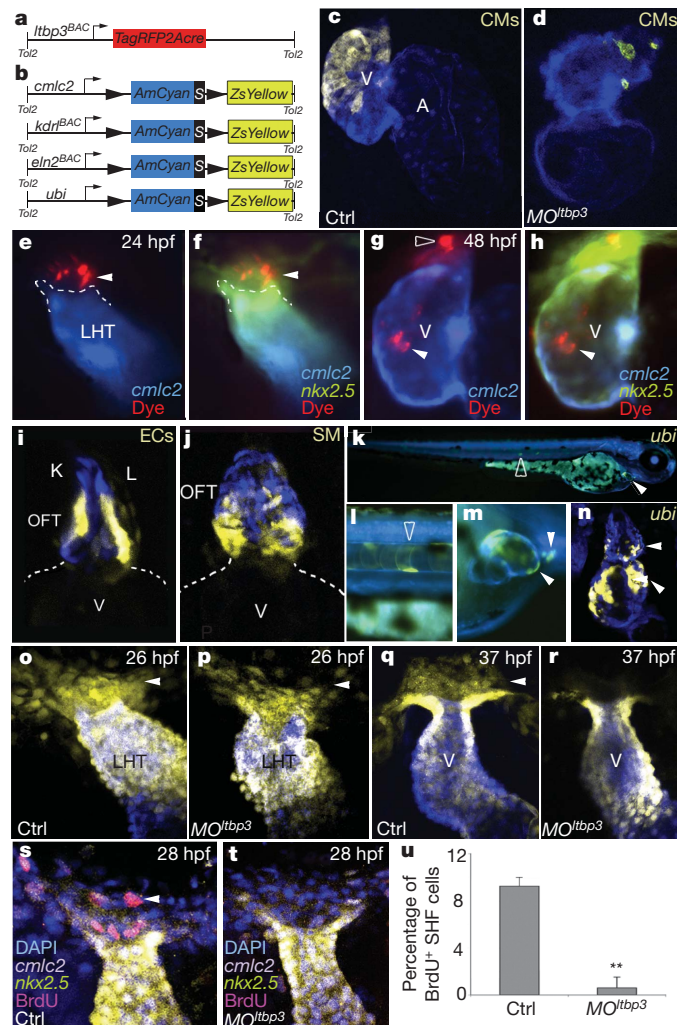
phenotype is remarkably similar to mouse anterior SHF mutants that die before septation with severe reductions in the primitive right ventricle and OFT^{15–17}. Therefore, we tested the hypothesis that *ltbp3*⁺ cells represent an SHF-like population that gives rise to some or all of the structures missing in *ltbp3* morphants. To that end, we used Cre/Lox-mediated lineage tracing to irreversibly mark *ltbp3*⁺ cells and their descendants that assume myocardial, endocardial/endothelial and smooth muscle cell fates. First, we derived a transgenic driver strain, *Tg(ltbp3::TagRFP2Acre)*, which co-expresses a red fluorescent protein (*TagRFP*) and Cre recombinase in *ltbp3*⁺ cells (Fig. 3a and Supplementary Fig. 7). Second, we generated three lineage-restricted reporter strains that carry a unique Cre-responsive ‘colour switching’ cassette (*AmCyan-Switch-ZsYellow* (CSY)) under transcriptional control of myocardial (*cmlc2*; also known as *myl7*), endothelial (*kdr1*), and smooth muscle precursor cell (*eln2*)^{7,14} promoters (Fig. 3b and Supplementary Fig. 8). We also generated a ubiquitous reporter strain using the zebrafish ubiquitin promoter¹⁸ (Fig. 3b and Supplementary Fig. 8). For each reporter, we confirmed Cre-dependent AmCyan to ZsYellow ‘colour switching’ caused by excision of the floxed AmCyan-STOP sequence from the reporter cassette (Supplementary Fig. 8).

Double transgenic progeny from the driver and myocardial reporter strains expressed ZsYellow protein in approximately the distal half of the ventricle (Fig. 3c) demonstrating that myocardium in this segment of the ventricle descends from *ltbp3*⁺ progenitors. Myocardial cells in the proximal OFT also arise from *ltbp3*⁺ cells (Supplementary Fig. 9). Furthermore, *ltbp3* morphant ventricles lacked *ltbp3*⁺-cell-derived cardiomyocytes, confirming that the distal ventricle is specifically affected in the morphants (Fig. 3d). In higher vertebrates, SHF cells labelled with a tracking dye after heart tube formation migrate into the OFT tract and ventricle^{12,19,20}. Similarly, we injected CellTrackerRed into the non-myocardial (*AmCyan*[−]), *nkx2.5*⁺ (ZsYellow⁺) region of double transgenic *Tg(nkx2.5::ZsYellow)*; *Tg(cmlc2::CSY)* embryos at the heart tube stage and observed subsequent dye migration into the distal ventricle and OFT (Fig. 3e–h). Taken together, these data demonstrate that *ltbp3*⁺ cells give rise to distal ventricular myocardium through late differentiation and accretion to the heart tube.

By crossing our driver line with the endothelial and smooth muscle reporter strains, we learned that *ltbp3*-expressing cells also give rise to endothelial and smooth muscle cells in the OFT (Fig. 3i, j). In smooth muscle reporter embryos, ZsYellow fluorescence was predominantly observed at the base of the OFT (Fig. 3j), indicating that *ltbp3*⁺ cells make a regionalized contribution to OFT smooth muscle. We did not observe colour switching in ventricular endocardium or pharyngeal arch artery endothelium, indicating that these cellular compartments are not derived from *ltbp3*⁺ cells (data not shown). Lastly, lineage tracing *ltbp3*⁺ cells with the ubiquitous reporter corroborated the conclusions drawn from lineage-restricted reporters (Fig. 3k–n). Taken together, these data demonstrate that *ltbp3*⁺ cells, in addition to giving rise to ventricular myocardium, also give rise to three cardiovascular lineages in the OFT.

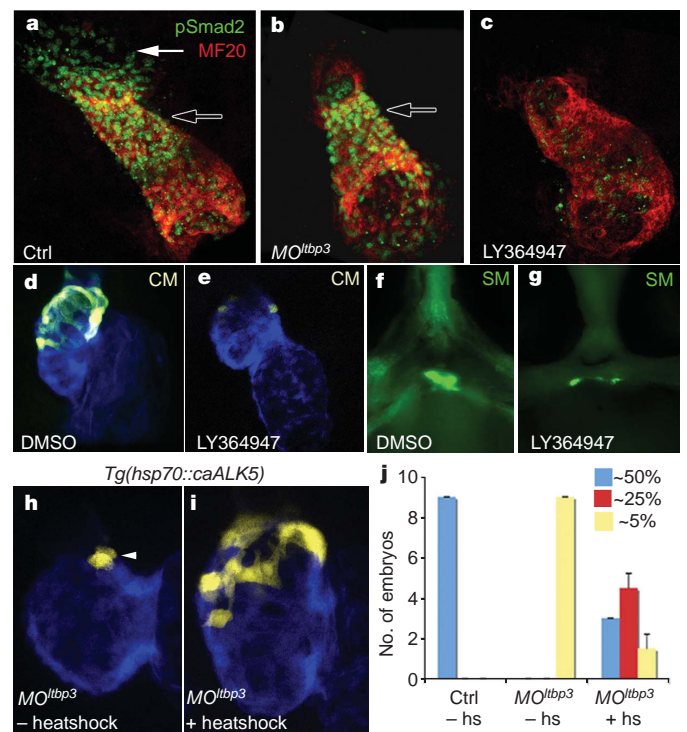
To elucidate the cellular mechanism(s) underlying the *ltbp3* morphant phenotype, we evaluated morphant embryos, soon after formation of the heart tube, for cellular defects in the extra-cardiac population of *ltbp3*⁺, *nkx2.5*⁺ cells at its outflow pole. Although *TagRFP*-expressing *ltbp3*⁺ cells are too faint to visualize in *Tg(ltbp3::TagRFP2Acre)* embryos at this developmental stage, non-myocardial ZsYellow-expressing *nkx2.5*⁺ cells are readily detected in *Tg(nkx2.5::ZsYellow)*; *Tg(cmlc2::CSY)* embryos (Fig. 1i–k). Shortly after heart tube formation, both control and morphant embryos harboured extra-cardiac *nkx2.5*⁺ cells, demonstrating that progenitor specification is not compromised by loss of *ltbp3* function (Fig. 3o, p). However, eleven hours later, this population was absent specifically in *ltbp3* morphants (Fig. 3q, r). The absence of TdT-mediated dUTP nick end labelling (TUNEL) indicated that progenitor survival is not compromised (data not shown). By contrast, BrdU staining revealed that progenitor proliferation is significantly reduced in *ltbp3* morphant embryos (Fig. 3s–u). Taken together, these

data support a model in which *ltbp3* functions in an autocrine fashion to stimulate proliferation or self-renewal of *ltbp3*⁺, *nkx2.5*⁺ progenitors to sustain adequate CPC numbers during their differentiation and



migration from the SHF. On the basis of this model, we speculate that knocking down *ltbp3* causes a premature depletion of CPCs from the SHF.

LTBP proteins are secreted with TGF-β ligands in a large latent complex that becomes anchored to the extracellular matrix before ligand activation¹¹. Because LTBP proteins are also required for secretion of TGF-β ligands, we used three approaches to test the hypothesis that defective TGF-β signalling underlies the cardiovascular phenotypes seen in *ltbp3* morphants. First, we evaluated the linear heart tube and surrounding regions for phosphorylated pSmad2 epitopes, a hallmark of active TGF-β signalling²¹. TGF-β signalling was observed in the heart tube proper and in extra-cardiac cells approximating the location of *ltbp3*⁺, *nkx2.5*⁺ progenitors (Fig. 4a). Loss of *ltbp3* function eliminated pSmad2 epitopes specifically in the extra-cardiac population, leaving signalling in the heart tube unaffected. These data demonstrate that *ltbp3* morphant embryos exhibit a highly localized defect in TGF-β signalling specifically in the SHF (Fig. 4b). Next, we evaluated whether small molecule inhibition of TGF-β signalling would phenocopy *ltbp3* morphants. To that end, we used a selective inhibitor (LY364947; ref. 22) of the TGF-β type I receptor ALK5 that effectively abrogated TGF-β signalling in zebrafish embryos (Fig. 4c). LY364947-treated embryos phenocopied *ltbp3* morphants, as evidenced by reductions in *ltbp3*⁺-cell-derived myocardium and OFT smooth muscle (Fig. 4d–g). Using LY364947, we also determined



the developmental time window, 26–36 h post-fertilization, during which TGF- β signalling is required for the formation of the distal ventricle (Supplementary Fig. 10), the same time window during which *nkx2.5*⁺ progenitors disappear in *ltbp3* morphants (Fig. 3o–r). Lastly, we attempted to rescue *ltbp3* morphant embryos by inducing ubiquitous expression of a constitutively active TGF- β type I receptor (constitutively active ALK5), the same receptor targeted by LY364947, predicted to be downstream of *ltbp3* and TGF- β ligands (Supplementary Fig. 10). In a majority of embryos, heat-shock inducible expression of constitutively active ALK5 partially or fully rescued the loss of distal myocardium seen in *ltbp3* morphants (Fig. 4h–j). Taken together, these results indicate that attenuation of TGF- β signalling underlies the cardiovascular phenotypes present in *ltbp3* morphant embryos.

Because *ltbp3*⁺ cells differentiate after formation of the heart tube in pharyngeal mesoderm, elongate the tube at the outflow pole, and give rise to three cardiovascular lineages in the OFT and myocardium in the distal ventricle, they show defining characteristics of the mouse anterior SHF. As such, *ltbp3*⁺ cells would also be predicted to encompass the zebrafish equivalents of the avian secondary and anterior heart fields^{1,19,23}.

Our findings are consistent with a recent study demonstrating that pre-existing ventricular cardiomyocytes do not give rise to new cardiomyocytes during elongation of the zebrafish heart tube⁹. The same authors report that the zebrafish homologue of *Islet1*, a well-known marker of mammalian SHF cells, is dispensable for formation of the distal ventricle, herein reported to derive from the zebrafish SHF. Thus, it seems that the critical role of *Islet1*, during anterior SHF development, has become minimized or replaced by another factor specifically in the fish lineage. Alternatively, *Islet1* function during mammalian anterior SHF development might have co-evolved with the four-chambered heart²⁴.

Although zebrafish do not have a right ventricle, *ltbp3*⁺ cells accrete the distal ventricular segment to the zebrafish heart tube much like SHF cells accrete the primitive right ventricle to the mouse heart tube before septation. In mouse, the SHF-derived segment becomes a distinct chamber after septation. In zebrafish, the SHF-derived segment instead augments the singular ventricular chamber, a trait that may have conferred an evolutionary advantage to a common ancestor of both zebrafish and mammals.

METHODS SUMMARY

Single and double, fluorescent and non-fluorescent, *in situ* hybridization and immunohistochemical stainings were performed using standard protocols. To analyse the *ltbp3* loss-of-function phenotype, we injected antisense *ltbp3* morpholinos into one-cell-stage wild-type and transgenic embryos. For genetic lineage tracing, a transgenic driver strain expressing Cre recombinase in *ltbp3*⁺ cells and four Cre-responsive 'colour switching' reporter strains were generated using standard methods. The driver strain was crossed individually to each of the reporter strains and their double transgenic progenies were analysed for ZsYellow protein fluorescence using confocal microscopy. To follow the migration of zebrafish SHF cells, a tracking dye was injected into the ZsYellow⁺, AmCyan[−] region of Tg(*nkx2.5::ZsYellow*); Tg(*cmhc2::CSY*) embryos at 24 h post-fertilization, the embryos were imaged immediately, and then again at 48 and/or 72 h post-fertilization. A transgenic strain carrying a cDNA encoding a constitutively active human TGF- β type I receptor (constitutively active ALK5) under control of the zebrafish heat shock promoter was generated and used to rescue the myocardial defect in *ltbp3* morphant embryos.

Received 21 October 2010; accepted 1 April 2011.

Published online 29 May 2011.

1. Dyer, L. A. & Kirby, M. L. The role of secondary heart field in cardiac development. *Dev. Biol.* **336**, 137–144 (2009).
2. Rochais, F., Mesbah, K. & Kelly, R. G. Signaling pathways controlling second heart field development. *Circ. Res.* **104**, 933–942 (2009).
3. Vincent, S. D. & Buckingham, M. E. How to make a heart: the origin and regulation of cardiac progenitor cells. *Curr. Top. Dev. Biol.* **90**, 1–41 (2010).

4. Olson, E. N. Gene regulatory networks in the evolution and development of the heart. *Science* **313**, 1922–1927 (2006).
5. Cai, C. L. *et al.* *Isl1* identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Dev. Cell* **5**, 877–889 (2003).
6. Meilhac, S. M. *et al.* The clonal origin of myocardial cells in different regions of the embryonic mouse heart. *Dev. Cell* **6**, 685–698 (2004).
7. Grimes, A. C. *et al.* Solving an enigma: arterial pole development in the zebrafish heart. *Dev. Biol.* **290**, 265–276 (2006).
8. Brade, T. *et al.* The amphibian second heart field: *Xenopus islet-1* is required for cardiovascular development. *Dev. Biol.* **311**, 297–310 (2007).
9. de Pater, E. *et al.* Distinct phases of cardiomyocyte differentiation regulate growth of the zebrafish heart. *Development* **136**, 1633–1641 (2009).
10. Grimes, A. C. *et al.* Phylogeny informs ontogeny: a proposed common theme in the arterial pole of the vertebrate heart. *Evol. Dev.* **12**, 552–567 (2010).
11. Rifkin, D. B. Latent transforming growth factor- β (TGF- β) binding proteins: orchestrators of TGF- β availability. *J. Biol. Chem.* **280**, 7409–7412 (2005).
12. Kelly, R. G., Brown, N. A. & Buckingham, M. E. The arterial pole of the mouse heart forms from *Fgf10*-expressing cells in pharyngeal mesoderm. *Dev. Cell* **1**, 435–440 (2001).
13. Chen, J. N. *et al.* Left-right pattern of cardiac BMP4 may drive asymmetry of the heart in zebrafish. *Development* **124**, 4373–4382 (1997).
14. Miao, M. *et al.* Differential expression of two tropoelastin genes in zebrafish. *Matrix Biol.* **26**, 115–124 (2007).
15. Ilagan, R. *et al.* *Fgf8* is required for anterior heart field development. *Development* **133**, 2435–2445 (2006).
16. Prall, O. W. *et al.* An *Nkx2-5/Bmp2/Smad1* negative feedback loop controls heart progenitor specification and proliferation. *Cell* **128**, 947–959 (2007).
17. von Both, I. *et al.* *Foxh1* is essential for development of the anterior heart field. *Dev. Cell* **7**, 331–345 (2004).
18. Mosimann, C. *et al.* Ubiquitous transgene expression and Cre-based recombination driven by the ubiquitin promoter in zebrafish. *Development* **138**, 169–177 (2010).
19. Waldo, K. L. *et al.* Conotruncal myocardium arises from a secondary heart field. *Development* **128**, 3179–3188 (2001).
20. Zaffran, S. *et al.* Right ventricular myocardium derives from the anterior heart field. *Circ. Res.* **95**, 261–268 (2004).
21. Todorovic, V. *et al.* Long form of latent TGF- β binding protein 1 (*Ltbp1L*) is essential for cardiac outflow tract septation and remodeling. *Development* **134**, 3723–3732 (2007).
22. Li, H. Y. *et al.* Dihydropyridopyrazole transforming growth factor- β type I receptor kinase domain inhibitors: a novel benzimidazole series with selectivity versus transforming growth factor- β type II receptor kinase and mixed lineage kinase-7. *J. Med. Chem.* **49**, 2138–2142 (2006).
23. Mjaatvedt, C. H. *et al.* The outflow tract of the heart is recruited from a novel heart-forming field. *Dev. Biol.* **238**, 97–109 (2001).
24. Kang, J. *et al.* *Isl1* is a direct transcriptional target of Forkhead transcription factors in second-heart-field-derived mesoderm. *Dev. Biol.* **334**, 513–522 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Whitman for advice on pSmad2 staining, D. Hami and M. Kirby for providing Eln2 antisera and their immunohistochemistry protocol, R. Cornell for providing *tAP2a* and *tAP2c* morpholinos, M. Whitman, A. Srinivasan, D. Langenau, E. Provost, S. Leach, R. Anderson, D. Stainier, I. Woods, and A. Schier for providing plasmids, J. W. Xiong for providing *clm*^{m39} fish, B. Barut and L. Zon for providing bacterial artificial chromosomes (BACs), and the MGH Nephrology Division for access to their confocal microscopy facilities. S.C., R.E.P. and W.H. were supported by NIH grant R01 ES012716 from the National Institute of Environmental Health Sciences. C.M. received support through an EMBO long-term fellowship and an HFSP long-term fellowship. This work was funded by the Cardiovascular Research Center at Massachusetts General Hospital, a Claflin Distinguished Scholar Award and Harvard Stem Cell Institute Seed Grant to C.E.B., and by awards from the National Heart Lung and Blood Institute (5R01HL096816), American Heart Association (Grant in Aid no. 10GRNT4270021), and Harvard Stem Cell Institute (Seed Grant) to C.G.B.

Author Contributions Y.Z. performed the majority of the experiments and analysed data; T.J.C., K.R.N., P.O., Y.L., A.G. and S.S. performed experiments and analysed data; C.M. provided the ubiquitin promoter prior to publication; S.A.C., R.E.P., W.H. and C.G.B. discovered that *ltbp3* transcripts are enriched in a cardiac fraction on day 3 post-fertilization; C.G.B. performed experiments including BAC recombineering; C.E.B. and C.G.B. co-directed the study, analysed data, and wrote the paper with input from all authors.

Author Information The zebrafish *ltbp3* cDNA sequence has been deposited into GenBank under accession number JF731042. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.E.B. (cburns6@partners.org) or C.G.B. (gcburns@cvc.harvard.edu).

Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins

Madeleine B. Borgia^{1,2}, Alessandro Borgia², Robert B. Best¹, Annette Steward¹, Daniel Nettels², Bengt Wunderlich², Benjamin Schuler² & Jane Clarke¹

A large range of debilitating medical conditions¹ is linked to protein misfolding, which may compete with productive folding particularly in proteins containing multiple domains². Seventy-five per cent of the eukaryotic proteome consists of multidomain proteins, yet it is not understood how interdomain misfolding is avoided. It has been proposed that maintaining low sequence identity between covalently linked domains is a mechanism to avoid misfolding³. Here we use single-molecule Förster resonance energy transfer^{4,5} to detect and quantify rare misfolding events in tandem immunoglobulin domains from the I band of titin under native conditions. About 5.5 per cent of molecules with identical domains misfold during refolding *in vitro* and form an unexpectedly stable state with an unfolding half-time of several days. Tandem arrays of immunoglobulin-like domains in humans show significantly lower sequence identity between neighbouring domains than between non-adjacent domains³. In particular, the sequence identity of neighbouring domains has been found to be preferentially below 40 per cent³. We observe no misfolding for a tandem of naturally neighbouring domains with low sequence identity (24 per cent), whereas misfolding occurs between domains that are 42 per cent identical. Coarse-grained molecular simulations predict the formation of domain-swapped structures that are in excellent agreement with the observed transfer efficiency of the misfolded species. We infer that the interactions underlying misfolding are very specific and result in a sequence-specific domain-swapping mechanism. Diversifying the sequence between neighbouring domains seems to be a successful evolutionary strategy to avoid misfolding in multidomain proteins.

Multidomain proteins comprise covalently linked, frequently similar domains, resulting in high effective local protein concentration. It is therefore probable that these proteins have evolved to avoid interdomain misfolding *in vivo*. Co-translational, 'domain-by-domain' folding is assumed to be important for avoiding misfolding⁶, but many multidomain proteins that are long lived or subject to tensile forces will fold and unfold numerous times during their lifetimes, resulting in neighbouring domains being unfolded at the same time, and therefore may be particularly vulnerable to misfolding. The giant muscle protein titin, for example, undergoes reversible domain unfolding that may have a role in muscle elasticity⁷.

Single-molecule techniques are ideal for detecting rare events⁸ such as misfolding in native conditions. Indeed, the first evidence for the misfolding of adjacent domains in long tandem arrays of the well-characterized twenty-seventh domain from the I band of titin, I27, was obtained using single-molecule atomic force microscopy⁹ (AFM). An alternative approach is single-molecule Förster resonance energy transfer^{4,5} (FRET), whose great sensitivity allows the detection of very small populations. FRET permits the mapping of intramolecular distances by means of the distance-dependent efficiency of excitation energy transfer between donor and acceptor fluorophores attached to specific positions of the protein¹⁰ (for details, see Supplementary Fig. 1).

We proposed that denaturation of tandem constructs of titin domains with guanidinium chloride (GdmCl), followed by rapid

refolding into native conditions, might allow formation of misfolded species, which should be detectable using single-molecule FRET¹¹. We labelled a tandem construct of I27 (I27–I27) in the A strand of domain 1 (E3C) and the G strand of domain 2 (N83C) (Fig. 1a) with a donor (Alexa Fluor 488) and an acceptor (Alexa Fluor 594) fluorophore, attached by means of cysteine residues engineered on the protein surface. For a misfolded domain formed by strands from domains 1 and 2, we predicted that these two strands must be adjacent for the domain to have the mechanical properties observed in the previous AFM experiments⁹. The correctly folded tandem domain would then have low transfer efficiency, whereas a misfolded state would have high transfer efficiency. A monomer of I27 was labelled in the corresponding positions to provide a model for the misfolded state (Fig. 1b). Labelling was found to have little effect on the stability of I27 (Supplementary Fig. 2), and doubly labelled proteins that had not previously been unfolded in GdmCl ('never unfolded') showed single, correctly folded populations with transfer efficiencies of $E = 0.37 \pm 0.01$ and 0.93 ± 0.01 for the tandem I27–I27 and monomer I27, respectively (Fig. 2a, b). (Uncertainties quoted in the text are s.d.)

We conducted refolding experiments by diluting unfolded I27–I27 into refolding buffer. The resulting transfer efficiency histograms then showed two populations (Fig. 2c): one corresponding to the correctly folded native state ($E = 0.37$) and one with precisely the same transfer efficiency as the analogously labelled monomer ($E = 0.93$). This

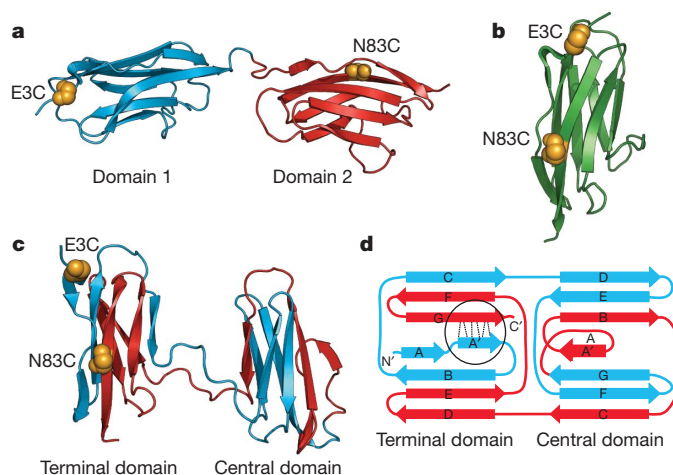


Figure 1 | Structures of native and misfolded I27 constructs. **a**, Natively folded I27–I27 tandem repeat with labelling positions highlighted (golden spheres). **b**, Native I27 crystal structure (Protein Data Bank ID, 1TIT). **c**, One of the domain-swapped, misfolded state structures formed in Go-model simulations. **d**, Schematic of the misfolded state topology in **c**: hydrogen bonds that are perpendicular to the direction of applied force in AFM mechanical unfolding are shown by dashed lines (circled). Four other misfolded state topologies were populated in the simulations (Supplementary Fig. 5b). We note that we cannot distinguish between such topologies from the results presented here.

¹University of Cambridge Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, UK. ²Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland.

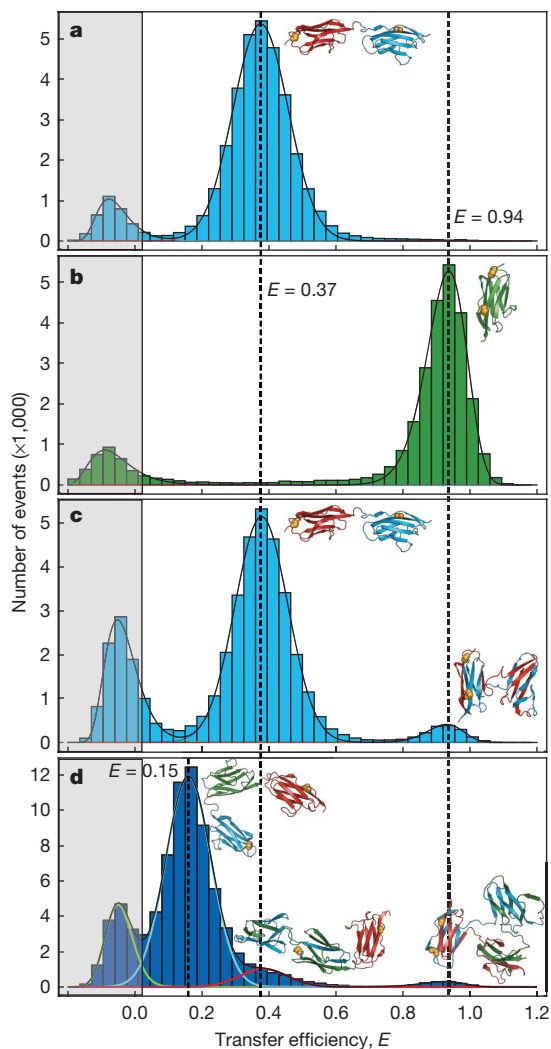


Figure 2 | Transfer efficiency histograms of doubly labelled I27 constructs.

a, Never-unfolded I27–I27. **b**, Never-unfolded monomeric I27. **c**, Refolded I27–I27. **d**, Refolded I27–I27–I27; fits of individual populations shown as coloured lines for clarity. Histograms are fitted with normal or log-normal distributions. The peak in the grey shaded area consists of events from molecules without an active acceptor fluorophore²⁸. We note that in these experiments a short, four-amino-acid linker (Arg-Ser-Glu-Leu) is included between the domains in the I27–I27 tandem to allow direct comparison with previous AFM and aggregation experiments^{3,9}. Never-unfolded I27–I27–I27 is shown in Supplementary Fig. 6.

observation reveals that the A strand of the first domain and the G strand of the second domain are arranged as in the monomer. A quantitative analysis (Supplementary Table 1) showed that $5.5 (\pm 0.2)\%$ of the molecules are found in the misfolded form.

On the basis of the results of the AFM studies⁹, we had supposed that the misfolded species consisted of a single strand-swapped titin domain with the remaining sequence unstructured and, thus, with an unfolding time similar to that of a native domain ($\tau \approx 34$ min; ref. 12). We therefore investigated the unfolding kinetics of the misfolded state¹³ (Fig. 3a). At high GdmCl concentrations, the decays in the number of high-transfer-efficiency events ($E > 0.8$), corresponding to the misfolded state, are fitted well by single exponentials (Fig. 3b) with rate constants slightly higher than the unfolding rate constants for wild-type I27 determined in ensemble measurements¹² (Fig. 3c). We can also estimate the unfolding rate constant of the correctly folded species, which agrees well with the ensemble data (Fig. 3c) (see Supplementary Fig. 3 and Supplementary Information). In the absence of denaturant, however, the misfolded species was surprisingly long lived,

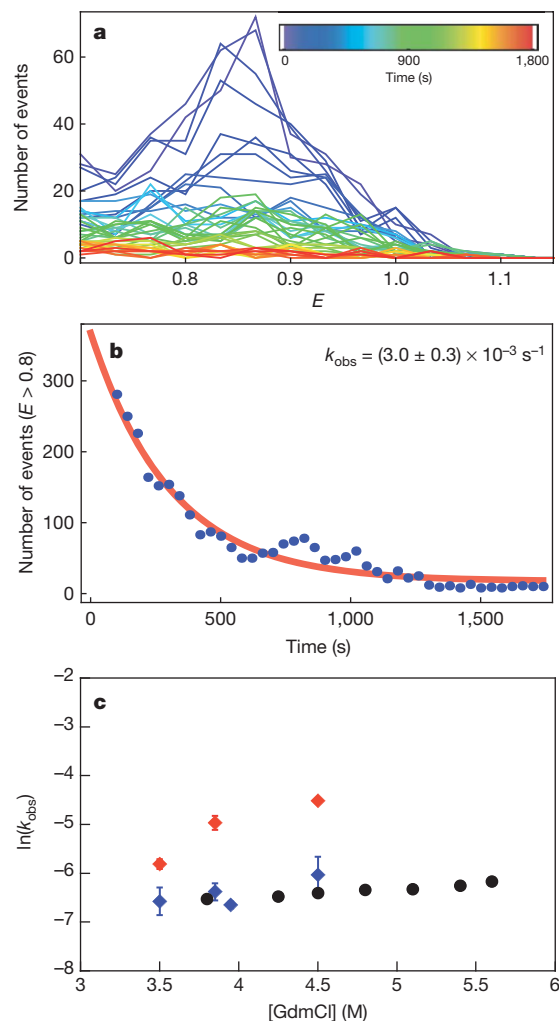


Figure 3 | Unfolding kinetics. **a**, Evolution of transfer efficiency histograms over time ($E \geq 0.7$) from single-molecule double-jump experiments in which refolded/misfolded I27–I27 (doubly labelled) was unfolded in 3.5 M GdmCl. Histograms were constructed for a moving window of 120 s that was shifted by 30 s for each increment (colour scale). **b**, The number of events with $E > 0.8$ for each histogram in **a** was summed and the resulting kinetics fitted with a single exponential decay. The observed rate constants (k_{obs}) are unaffected by different window sizes or the use of non-overlapping windows. **c**, Unfolding rate constants for the wild-type I27 monomer (black; ensemble data from ref. 12) and for the misfolded and natively folded states of I27–I27 from single-molecule measurements (red and blue, respectively), as functions of GdmCl concentration. Each point was obtained by fitting data sets composed by merging at least six repeats, with error bars representing the standard error of the fit (Methods). For some data points, the error bars are smaller than the symbols. I27 domains have the same unfolding rate constants in tandem repeat proteins as in isolated domains.

converting to the correctly folded form only on a timescale of days (Supplementary Fig. 4). The formation of the misfolded structure is thus under kinetic, rather than thermodynamic, control. Its remarkable kinetic stability clearly distinguishes the misfolded species described here from short-lived, partly folded intermediates^{14–16} sometimes termed ‘misfolded’ because they contain some non-native interactions¹⁷.

An explanation for the slow unfolding under native conditions is suggested by folding simulations of I27–I27 with a Gō-like model¹⁸. In these simulations, only native interactions are attractive, and interactions between a given pair of residues are considered equal, independently of whether they are in the same or different domains. Although most trajectories result in two correctly folded domains, misfolded species with two fully folded, strand-swapped domains are occasionally formed. Five different strand-swapped topologies were observed

(Fig. 1c, d and Supplementary Fig. 5). Such an extensively misfolded structure explains its persistence: correct folding cannot occur while either misfolded domain remains folded. Because refolding rate constants are much higher than unfolding rate constants under native conditions, the simultaneous unfolding of both domains is very unlikely and conversion to the native state is extremely slow¹⁹.

We can test the validity of our model further by investigating the refolding of a three-domain tandem of I27 with the FRET labels in domain 1 (E3C) and domain 3 (N83C). If domain-swapped structures were to be formed, we would expect to see two misfolded populations: one with the FRET efficiency of the monomer (misfolding between domains 1 and 3) and another population with the efficiency of the I27–I27 tandem (misfolding between domains 1 and 2 or 2 and 3). This is precisely what we observe (Fig. 2d). The proportion of monomer-like (high-FRET) species in the trimeric tandem is significantly lower than before ($2.8 (\pm 0.6)\%$); this is likely to reflect the lower probability of association between domains that are more distant in sequence. The population of the misfolded species with dimer-like FRET efficiency ($9 (\pm 2)\%$) was instead almost twice as large as in the two-domain tandem ($5.5 (\pm 0.2)\%$); this is probably due to the two alternative possibilities to misfold in an analogous way to the dimeric tandem (domain 1 with 2 and domain 2 with 3). Simulations with the Gō-like model also predict domain-swapped structures with monomer- and dimer-like FRET efficiencies, with relative populations similar to experiment (Supplementary Fig. 6).

Much work has been dedicated to investigating the sequence specificity of protein aggregation^{20,21}, including the hypothesis that there is selective pressure to prevent oligomerization by a domain-swapping mechanism^{22,23}. Misfolding is often considered to precede aggregation, suggesting that sequence-specific behaviour observed in aggregation also applies to misfolding. Our single-molecule FRET experiments allow us to test this hypothesis directly by investigating mixed tandem constructs, I27–I28 and I27–I32 (Fig. 4). Indeed, the construct comprising I27 and I28, natural neighbours in titin with only 24% sequence identity, did not yield any detectable population of high-FRET misfolded species on refolding (Fig. 4c, d). However, misfolding is seen on refolding of I27–I32 (sequence identity, 42%; Fig. 4e, f), to the same extent as for I27–I27 (Supplementary Table 1). This misfolded species also unfolds with the same rate constant as I27–I27 (Supplementary Fig. 7). The I27–I32 misfold is consistent with previous experiments showing chimaeric domains of I27 and I32 to be stable^{24,25}. This result strongly supports the idea that protein misfolding is sequence specific. In proteins where sequence identity between neighbouring domains is high, the topology may prevent formation of stable misfolded species¹⁹.

Misfolding in our experiments is more frequent than had been observed in AFM experiments⁹, suggesting that the tethering in those experiments reduces misfolding; this might be advantageous for titin domains *in vivo*. Unfolding of the misfolded species observed with AFM showed them to have the same mechanical resistance as correctly folded I27, but with twice the chain length being released on unfolding. Our results are entirely compatible with this finding. Although the misfolded species has two folded domains, and is therefore stable in folding conditions, only the terminal domain would experience shearing of the hydrogen bonds between the parallel A' and G strands (Fig. 1d, circled) perpendicular to the direction of the applied force, in the same way as the correctly folded I27²⁶, resulting in the same mechanical stability. Because force is not applied to the A and G strands in the central domain, this domain is likely to unfold at low force, together with the terminal domain. This hypothesis is supported by simulations (Supplementary Fig. 8).

Our results suggest that diversifying the sequence composition between neighbouring domains is an effective evolutionary strategy to ensure efficient folding in multidomain proteins and avoid the formation of stable misfolded species. This adds a significant piece to the puzzle of understanding the problems encountered during the crucial evolutionary transition from single to multidomain proteins.

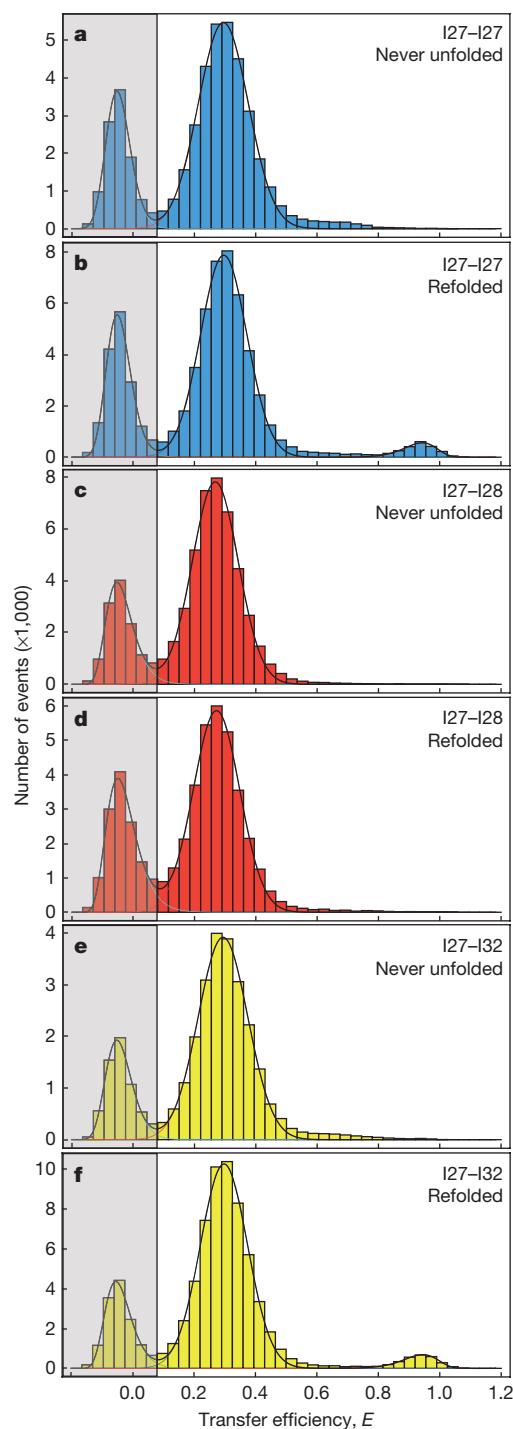


Figure 4 | Transfer efficiency histograms of tandem constructs with identical and non-identical domains. **a, b**, I27–I27 never-unfolded control (**a**) and refolded (**b**) constructs. **c, d**, I27–I28 never-unfolded control (**c**) and refolded (**d**) constructs. **e, f**, I27–I32 never-unfolded (**e**) control and refolded (**f**) constructs. To mimic the natural protein, no linker was added between the domains in these experiments. We note that the frequency of misfolding was the same for I27–I27 with and without the linker, $5.5 (\pm 0.2)\%$ and $5.7 (\pm 0.5)\%$, respectively (compare Fig. 2c with Fig. 4b and Supplementary Table 1). Addition of the four-amino-acid linker also made no difference to the results for I27–I28 (Supplementary Fig. 9).

METHODS SUMMARY

For details of protein production, ensemble equilibrium measurements and labeling, see Methods. Single-molecule experiments, instrumentation, data reduction and analysis are also detailed in Methods. The resulting relative populations from all experiments and analysis techniques are summarized in Supplementary Table 1.

Folding simulations using a Gō-like model were run using the CHARMM code²⁷ as described in Methods. For details of mechanical unfolding simulations, see Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 January; accepted 1 April 2011.

Published online 29 May 2011.

1. Gregersen, N., Bross, P., Vang, S. & Christensen, J. H. Protein misfolding and human disease. *Annu. Rev. Genomics Hum. Genet.* **7**, 103–124 (2006).
2. Jaenicke, R. & Seckler, R. Protein misassembly *in vitro*. *Adv. Protein Chem.* **50**, 1–59 (1997).
3. Wright, C. F., Teichmann, S. A., Clarke, J. & Dobson, C. M. The importance of sequence diversity in the aggregation and evolution of proteins. *Nature* **438**, 878–881 (2005).
4. Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C. & Ha, T. Advances in single-molecule fluorescence methods for molecular biology. *Annu. Rev. Biochem.* **77**, 51–76 (2008).
5. Schuler, B. & Eaton, W. A. Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* **18**, 16–26 (2008).
6. Fedorov, A. N. & Baldwin, T. O. Cotranslational protein folding. *J. Biol. Chem.* **272**, 32715–32718 (1997).
7. Li, H. *et al.* Reverse engineering of the giant muscle protein titin. *Nature* **418**, 998–1002 (2002).
8. Borgia, A., Williams, P. M. & Clarke, J. Single-molecule studies of protein folding. *Annu. Rev. Biochem.* **77**, 101–125 (2008).
9. Oberhauser, A. F., Marszalek, P. E., Carrion-Vasquez, M. & Fernandez, J. M. Single protein misfolding events captured by atomic force microscopy. *Nature Struct. Biol.* **6**, 1025–1028 (1999).
10. Stryer, L. Fluorescence energy transfer as a spectroscopic ruler. *Annu. Rev. Biochem.* **47**, 819–846 (1978).
11. Gambin, Y. *et al.* Direct single-molecule observation of a protein living in two opposed native structures. *Proc. Natl Acad. Sci. USA* **106**, 10153–10158 (2009) CrossRef.
12. Fowler, S. B. & Clarke, J. Mapping the folding pathway of an immunoglobulin domain: structural detail from phi value analysis and movement of the transition state. *Structure* **9**, 355–366 (2001).
13. Hofmann, H. *et al.* Single-molecule spectroscopy of protein folding in a chaperonin cage. *Proc. Natl Acad. Sci. USA* **107**, 11793–11798 (2010).
14. Ivarsson, Y., Travaglini-Alloccatelli, C., Brunori, M. & Gianni, S. Folding and misfolding in a naturally occurring circularly permuted PDZ domain. *J. Biol. Chem.* **283**, 8954–8960 (2008).
15. Gianni, S. *et al.* Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. *Nature Struct. Mol. Biol.* **17**, 1431–1437 (2010).
16. Korzhnev, D. M., Religa, T. L., Banachewicz, W., Fersht, A. R. & Kay, L. E. A transient and low-populated protein-folding intermediate at atomic resolution. *Science* **329**, 1312–1316 (2010).
17. Capaldi, A. P., Kleanthous, C. & Radford, S. E. Im7 folding mechanism: misfolding on a path to the native state. *Nature Struct. Biol.* **9**, 209–216 (2002).
18. Yang, S. *et al.* Domain swapping is a consequence of minimal frustration. *Proc. Natl Acad. Sci. USA* **101**, 13786–13791 (2004).
19. Arora, P., Hammes, G. G. & Oas, T. G. Folding mechanism of a multiple independently-folding domain protein: double B domain of protein A. *Biochemistry* **45**, 12312–12324 (2006).
20. Jaenicke, R. Folding and association of proteins. *Prog. Biophys. Mol. Biol.* **49**, 117–237 (1987).
21. Straub, J. E. & Thirumalai, D. Principles governing oligomer formation in amyloidogenic peptides. *Curr. Opin. Struct. Biol.* **20**, 187–195 (2010).
22. Bennett, M. J., Sawaya, M. R. & Eisenberg, D. Deposition diseases and 3D domain swapping. *Structure* **14**, 811–824 (2006).
23. Mitraki, A. Protein aggregation from inclusion bodies to amyloid and biomaterials. *Adv. Protein Chem. Struct. Biol.* **79**, 89–125 (2010).
24. Borgia, A., Steward, A. & Clarke, J. An effective strategy for the design of proteins with enhanced mechanical stability. *Angew. Chem. Int. Ed.* **47**, 6900–6903 (2008).
25. Balamurali, M. M. *et al.* Recombination of protein fragments: a promising approach toward engineering proteins with novel nanomechanical properties. *Protein Sci.* **17**, 1815–1826 (2008).
26. Lu, H., Israilewitz, B., Krammer, A., Vogel, V. & Schulten, K. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.* **75**, 662–671 (1998).
27. Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).
28. Dahan, M. *et al.* Ratiometric measurement and identification of single diffusing molecules. *Chem. Phys.* **247**, 85–106 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Wellcome Trust (grant number, 064417), the Swiss National Science Foundation (to B.S.) and the Swiss National Center of Competence in Research in Structural Biology (to B.S.). M.B.B. was supported by a UK Medical Research Council studentship. A.B. is supported by a Marie Curie Intra-European Fellowship. R.B.B. is supported by a Royal Society University Research Fellowship. J.C. is a Wellcome Trust Senior Research Fellow. We thank H. Hofmann, A. Soranno and A. Hoffmann for discussions and contributions to data analysis.

Author Contributions M.B.B., A.B., B.S. and J.C. designed the investigation. M.B.B. and A.B. performed the experiments. R.B.B. performed the simulations. D.N. and B.W. built the single-molecule instrumentation. D.N. provided data analysis software. A.S. cloned the gene of the trimeric tandem construct. M.B.B. performed the analysis. M.B.B., J.C. and B.S. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.C. (jc162@cam.ac.uk) or B.S. (schuler@bioc.uzh.ch).

METHODS

Protein expression and labelling. Cysteine residues were introduced by site-directed mutagenesis: for the two-domain constructs, E3C in domain 1 (always I27), N83C in domain 2 if I27 or I32, and K83C in domain 2 if I28 (with the numbering relative to a single domain); and for the three-domain construct of I27, E3C in domain 1 and N83C in domain 3. DNA sequencing confirmed the mutagenesis.

I27 monomer and the I27–I27, I27–I27–I27, I27–I28 and I27–I32 tandems, with the engineered surface cysteines, were expressed as described previously^{29,30}. Labelling was carried out using Alexa Fluor 488 (donor) and Alexa Fluor 594 (acceptor) maleimides (Invitrogen) according to the manufacturer's procedures. The dyes were mixed simultaneously with reduced protein in equimolar ratios and incubated at 4 °C for ~10 h. Unreacted dye was removed by gel filtration and the differently labelled variants were separated by ion-exchange chromatography (MonoQ 5/50 GL; GE Healthcare Biosciences AB). I27 has two intrinsic cysteines that were not removed as they are buried in the native state and all labelling was carried out on folded protein in native conditions.

Ensemble measurements. Equilibrium measurements were performed for the doubly labelled I27 monomer to check the effect of labelling (Supplementary Fig. 2). Experiments were performed in GdmCl on a Cary Eclipse fluorimeter (Varian Inc.) monitoring intrinsic tryptophan fluorescence as described previously²⁹, but with lower protein concentrations (0.05–0.5 µM) and the addition of 0.001% Tween 20.

Single-molecule instrumentation. Observations of single-molecule fluorescence were made using a custom-built confocal microscope equipped with a continuous-wave, 488-nm solid-state laser (FCD488-010, JDSU) and an Olympus UplanApo ×60/1.20W objective. After passage through a dichroic mirror that separates excitation and emission light (500DCXR, Chroma Technology), fluorescence emission passed through a 100-µm pinhole and was split into donor and acceptor fluorescence by a second dichroic mirror (585DCXR, Chroma Technology). Donor fluorescence then passed a filter (ET525/50M, Chroma Technology) before being focused onto a single-photon avalanche diode (MPD 100ct, Micro Photon Devices). Similarly, acceptor fluorescence passed a filter (QT 650/100) before being focused onto a single-photon avalanche diode (SPCM-AQR-13, PerkinElmer Optoelectronics). The arrival time of every photon was recorded with a two-channel, time-correlated, single-photon counting module (PicoHarp300, PicoQuant). All measurements were performed with a laser power of 100 µW, measured at the back aperture of the objective (beam waist, 8 mm).

Single-molecule equilibrium measurements. All experiments were performed at protein concentrations between 0.5 and 25 pM, in the same final solution conditions: PBS; 0.001% Tween 20; 140 mM β-mercaptoethanol; 20 mM cysteamine hydrochloride (single-molecule buffer). Tween 20 (Pierce) was used to prevent surface adhesion of the proteins³¹, whereas the photoprotective agents β-mercaptoethanol (Sigma) and cysteamine hydrochloride (Sigma) were used to minimize chromophore damage³². Experiments on never-unfolded proteins were conducted by mixing protein in PBS (0.01% Tween, 10 mM β-mercaptoethanol) 1:99 with 0.04 M GdmCl in single-molecule buffer (to mimic the final conditions in the refolding experiments). Refolding experiments were performed by mixing protein unfolded in 4.4 M GdmCl (PBS, 0.01% Tween, 10 mM β-mercaptoethanol) 1:100 with single-molecule buffer, to a final GdmCl concentration of 0.04 M. Measurements over 8 to 10 h were made for all constructs with one or more repeats. The absence of aggregates was ensured in all experimental conditions as previously described³³. Fluorescence bursts were identified by combining successive photons separated by 150 µs or less, and events comprising 35 or more photons were kept for analysis. Transfer efficiencies were corrected for quantum yields, cross-talk and direct excitation as described previously^{34,35}.

Populations of correctly folded and misfolded molecules in the transfer efficiency histograms were quantified using two methods. Where possible, peaks were fitted using a Gaussian distribution (for populations where $0.1 \leq E \leq 0.8$) or a log-normal distribution (for populations where $E \geq 0.8$), and the resulting fits integrated. Transfer efficiencies quoted in the main text denote the average E value obtained from these fits, with standard deviations calculated from multiple experiments. The populations of misfolded species were determined relative to the sum of natively folded and misfolded populations. Alternatively, ranges of transfer efficiencies were chosen for each population, the corresponding number of bursts were summed and the relative populations of misfolded species were computed.

Only the latter method could be used in the 'never-unfolded' control measurements and all experiments involving I27–I28 tandems, as no misfolded population was observed. The resulting relative populations from all experiments and analysis techniques are summarized in Supplementary Table 1.

Single-molecule kinetic measurements. Unfolding of the misfolded species was achieved by mixing a previously refolded protein sample (prepared as for refolding experiments above, but with higher protein concentration) with GdmCl in single-molecule buffer. To obtain sufficient statistics, at least six repeat measurements were made for each GdmCl concentration and the resulting photon trajectories were merged to give one data set. A moving-window analysis¹³ was applied to the merged, time-resolved photon trajectory with a window size (Δt) of 120 s. Transfer efficiency histograms were calculated from the bursts in that time window, and the window was shifted by $\Delta t/3$ to form each successive time point (Fig. 3a). A time $t = t_s + \Delta t/2$, where t_s is the start time of the window, was assigned to each histogram, and the number of events with $E \geq 0.8$ in each histogram as a function of time was fitted to a single exponential (Fig. 3b). The resulting rate constants were robust for different window sizes and non-overlapping windows. Standard errors were taken from the covariance matrix of the fit (weighted by the average inverse variance of the residuals of the data points with respect to an unweighted fit).

The analysis of the unfolding of the native state from the same type of experiment is detailed in Supplementary Information and Supplementary Fig. 5.

Simulations. A coarse-grained Gō-like model was generated on the basis of the structure of I27 (Protein Data Bank ID, 1TIT³⁶) using a standard procedure³⁷. Briefly, all bond lengths are fixed by constraints, harmonic terms are used for the angles, a knowledge-based potential is used for the torsion angles and non-bonded interactions are treated with a Gō-like potential in which only interactions formed between residues in the folded protein are attractive (with relative strengths given by the Miyazawa–Jernigan matrix), all others being repulsive. Two or three identical I27 sequences were linked by four-residue repulsive linkers to treat the two and three domain tandems, respectively. Interactions between residues i and j in different domains were treated exactly like the interactions between those residues in the same domain, and interactions with the linker were repulsive. A simulation temperature was chosen such that the folding barrier was at least $3k_B T$, by using as a lower bound the free-energy barrier projected onto the fraction of native contacts. Folding simulations were run, starting from fully extended configurations, using Langevin dynamics with a friction of 0.1 ps^{-1} and a time step of 10 fs. The final structures were clustered using a simple leader–follower algorithm with a cut-off of 0.15 on the r.m.s. distance between contact maps.

Mechanical unfolding simulations were performed in which a force of 150 pN was applied to the ends of the I27–I27 tandem, starting from structures belonging to either the folded cluster or one of the misfolded clusters, and monitoring unfolding by measuring the end–end distance. The CHARMM molecular simulation package was used for all calculations²⁷.

29. Scott, K. A., Steward, A., Fowler, S. B. & Clarke, J. Titin; a multidomain protein that behaves as the sum of its parts. *J. Mol. Biol.* **315**, 819–829 (2002).
30. Steward, A., Toca-Herrera, J. L. & Clarke, J. Versatile cloning system for construction of multimeric proteins for use in atomic force microscopy. *Protein Sci.* **11**, 2179–2183 (2002).
31. Schuler, B., Lipman, E. A. & Eaton, W. A. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **419**, 743–747 (2002).
32. Nettels, D. *et al.* Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proc. Natl Acad. Sci. USA* **106**, 20740–20745 (2009).
33. Hillger, F., Nettels, D., Dorsch, S. & Schuler, B. Detection and analysis of protein aggregation with confocal single molecule fluorescence spectroscopy. *J. Fluoresc.* **17**, 759–765 (2007).
34. Schuler, B. Application of single molecule Förster resonance energy transfer to protein folding. *Protein Folding Protocols* (eds Bai, Y. & Nussinov, R.) 115–138 (Humana Press, 2007).
35. Hoffmann, A. *et al.* Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc. Natl Acad. Sci. USA* **104**, 105–110 (2007).
36. Imptora, S., Politou, A. S. & Pastore, A. Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. *Structure* **4**, 323–337 (1996).
37. Karanicolas, J. & Brooks, C. L. III. Improved Gō-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. *J. Mol. Biol.* **334**, 309–325 (2003).

A gene regulatory network controlling the embryonic specification of endoderm

Isabelle S. Peter¹ & Eric H. Davidson¹

Specification of endoderm is the prerequisite for gut formation in the embryogenesis of bilaterian organisms. Modern lineage labelling studies^{1–3} have shown that in the sea urchin embryo model system, descendants of the veg1 and veg2 cell lineages produce the endoderm, and that the veg2 lineage also gives rise to mesodermal cell types. It is known that Wnt/ β -catenin signalling is required for endoderm specification^{4–6} and Delta/Notch signalling is required for mesoderm specification^{7–9}. Some direct *cis*-regulatory targets of these signals have been found^{10,11} and various phenomenological patterns of gene expression have been observed in the pre-gastrular endoderm. However, no comprehensive, causal explanation of endoderm specification has been conceived for sea urchins, nor for any other deuterostome. Here we propose a model, on the basis of the underlying genomic control system, that provides such an explanation, built at several levels of biological organization. The hardwired core of the control system consists of the *cis*-regulatory apparatus of endodermal regulatory genes, which determine the relationship between the inputs to which these genes are exposed and their outputs. The architecture of the network circuitry controlling the dynamic process of endoderm specification then explains, at the system level, a sequence of developmental logic operations, which generate the biological process. The control system initiates non-interacting endodermal and mesodermal gene regulatory networks in veg2-derived cells and extinguishes the endodermal gene regulatory network in mesodermal precursors. It also generates a cross-regulatory network that specifies future anterior endoderm in veg2 descendants and institutes a distinct network specifying posterior endoderm in veg1-derived cells. The network model provides an explanatory framework that relates endoderm specification to the genomic regulatory code.

Transcription factors, which are the products of regulatory genes, implement the genomic code for development by determining the set of expressed genes, and thus biological function. The spatially restricted expression of regulatory genes produces specific combinations of transcription factors, or regulatory states, in distinct spatial domains of the embryo. The complete set of regulatory interactions required for the formation and propagation of regulatory states explains the process of developmental specification, and this explanation is the ultimate goal of gene regulatory network (GRN) analysis.

We have systematically analysed the GRN that determines the specification of the future endoderm in the embryo of the sea urchin *Strongylocentrotus purpuratus*, up to gastrulation. In this embryo, endoderm is derived from two cell lineages, which arise by a canonical and invariant cleavage process (Supplementary Fig. 1). The anterior compartments of the gut are formed by cells of the veg2 lineage, which is also the progenitor lineage of most mesodermal cell types. The posterior endoderm is formed by derivatives of the veg1 lineage. Comprehensive surveys^{12–16} of all predicted transcription factors in this genome showed that 14 regulatory genes are expressed specifically in endoderm-precursor cells before the beginning of gastrulation (30 h post-fertilization). Spatial expression patterns for these genes are summarized in Fig. 1a–c on the basis of evidence from

double-fluorescent *in situ* hybridization (DFISH; Supplementary Figs 2 and 3) and earlier reports. We also provide a comprehensive digital summary of expression patterns at 3-h intervals for these and many additional genes in Supplementary Fig. 4.

To establish a causal explanation for the dynamic process of regulatory-state separation in the respective spatial fate domains of the veg2 and veg1 lineages, we carried out a system-wide perturbation analysis (more than 6,500 data points; Supplementary Fig. 5). The expression of each transcription factor was downregulated by treating embryos with specific morpholino antisense oligonucleotides (MASO), and the effects on all other regulatory genes, as well as on many representative genes expressed in non-endodermal domains, were measured quantitatively and often assessed spatially as well. These perturbation results were interpreted using the logic and evidence detailed in Supplementary Tables 1–3. The probable direct interactions, some of which have already been confirmed by *cis*-regulatory analysis, are represented in Fig. 1d–f. The spatial regulatory-state matrix in Fig. 1a–c can be considered as the output of the gene interaction matrix in Fig. 1d–f. Perturbation results were combined with previous *cis*-regulatory evidence to formulate the GRN model.

By 15 h post-fertilization, endoderm progenitor cells constitute two distinct, concentrically arranged regulatory states (Supplementary Fig. 1)¹⁷. In the more vegetal tiers of cells, encompassing the veg2 endoderm precursor cells, eight regulatory genes are rapidly turned on. Most of these are under the spatial control of *cis*-regulatory Tcf sites^{11,17–19}, which bind the factor mediating Wnt signal transduction, and there are a few additional regulatory interactions among them¹⁷. By contrast, only one regulatory gene, *even skipped* (*eve*), is expressed in the peripheral veg1 endoderm precursors, with no detectable impact on any other regulatory gene at this stage¹⁷. Before 18 h after fertilization, the two endoderm regulatory states are expressed in most or all of the cells in the veg1 and veg2 lineages.

At this stage, the veg2 lineage consists of two concentric rings of cells, the inner ring destined to become mesoderm and the outer ring, anterior endoderm (Supplementary Fig. 1). The future mesoderm expresses both endoderm and mesoderm GRNs, whereas the future endoderm expresses only an endoderm regulatory state¹⁷. Thus, DFISH using probes that detect the endoderm regulatory gene *forkhead box A* (*foxa*) and the mesoderm regulatory gene *glial cells missing* (*gcm*) shows that there is overlapping expression of both genes in an inner ring of veg2-derived cells, whereas the peripheral cells of the veg2 lineage (the presumptive endoderm) express *foxa* alone (Fig. 2a). The *gcm* gene is at the top of the early mesoderm GRN hierarchy and is activated by signalling through the Delta/Notch pathway, via its *cis*-regulatory Suppressor of Hairless (Su(H)) target sites¹⁰. The expression of *gcm* is therefore restricted to the inner ring of veg2-derived cells, which are exposed to the Delta-presenting skeletogenic cells at the vegetal pole (Supplementary Fig. 1). *Cis*-regulatory modules that respond to Tcf or Su(H) act as 'X,1–X' spatial information processors²⁰ in that they stimulate or permit gene expression in cells ('X') with an activated signal transduction pathway, but repress the same target genes in all other cells ('1–X'). Thus, Su(H) and Tcf *cis*-regulatory

¹Division of Biology, California Institute of Technology, Pasadena, California 91125, USA.

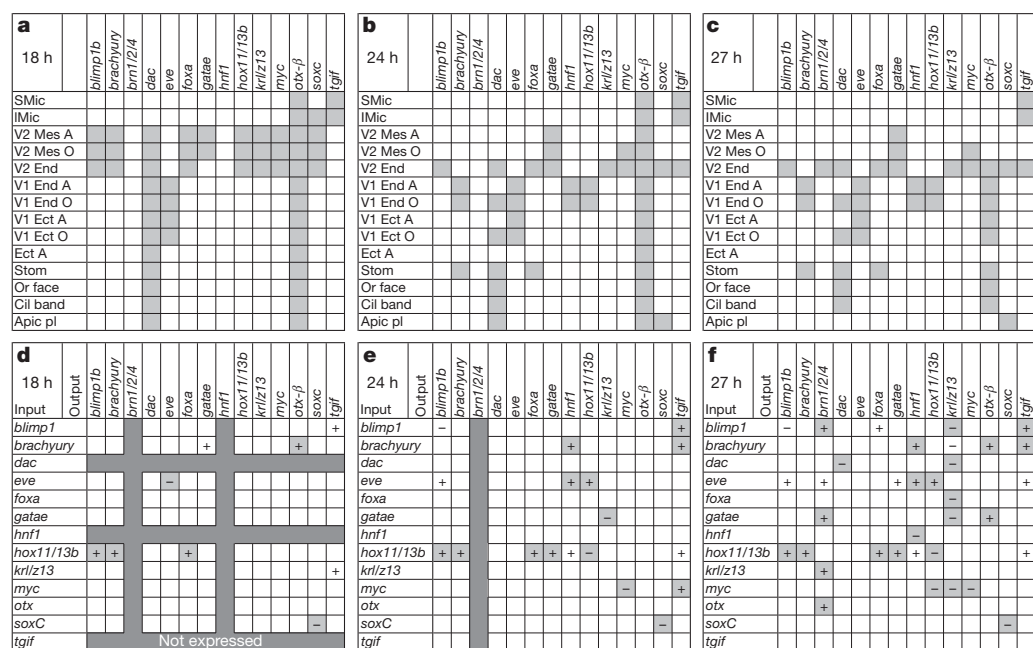


Figure 1 | Endodermal gene expression and perturbation matrix. **a–c**, Spatial expression (grey cells) of 14 endodermal regulatory genes at three time points post-fertilization. A, aboral; Apic pl, apical plate; Cil band, ciliated band; Ect, ectoderm; End, endoderm; LMic, large micromeres; Mes, mesoderm; O, oral; Or face, oral face; SMic, small micromeres; Stom, stomodeum; V1, veg1; V2, veg2. **d–f** Interactions among regulatory genes at three time points post-fertilization (data from ref. 17 and Supplementary Fig. 5). The change in output-gene expression after injection of a morpholino oligonucleotide

targeting the input gene is denoted ‘–’ if expression is significantly increased (that is, the input gene represses the output gene) and ‘+’ if expression is significantly decreased (the input gene activates the output gene). Changes incorporated as regulatory linkages in the endoderm GRN are indicated by light grey cells (Supplementary Table 1); white cells with ‘+’ or ‘–’ denote significant effects that are not considered to be direct (Supplementary Tables 2 and 3); dark grey cells denote genes that are not expressed at the time point shown.

interactions account for the spatial specificity of the initial, co-existing GRNs in the veg2 lineage.

Within a few hours, the sea urchin embryo accomplishes one of the most important regulatory transitions in embryonic development, the permanent separation of endodermal fate from mesodermal fate in sister cells descendant from the same endomesodermal precursors. All but one

of the eight endodermal regulatory genes cease to be expressed in mesodermal precursors by 24 h post-fertilization, the exception being the *myc* gene (Supplementary Fig. 2). Thus, the expression domains of *foxa* and *gcm*, which are partially overlapping at 16 h, become exclusive after 18 h, as shown by DFISH (Fig. 2a, b). The genomic mechanism of regulatory-state exclusion is elegant: the same Tcf sites that are used to initiate the endoderm GRN in the veg2 lineage are used again to extinguish it in mesoderm precursors. The mechanism depends on Delta/Notch signaling, which is also the inducer of mesoderm gene expression. In embryos with perturbed expression of either Delta or Notch, the endodermal regulatory genes *foxa*, *blimp1b* and *dachshund* (*dac*) continue to be expressed in the presumptive mesodermal domain at 24 h (Fig. 2c, d, Supplementary Fig. 6 and data not shown). A similar result was reported for *foxa* in another sea urchin species²¹. The exclusion of the endoderm GRN is independent of mesoderm specification per se, as perturbation of *gcm* expression does not lead to ectopic *foxa* expression in mesoderm progenitors (Fig. 2e, f). A *foxa* cis-regulatory study has demonstrated directly that Tcf target sites are required for transcriptional repression of this gene in mesoderm precursor cells¹¹. A possible explanation is that in cells receiving Notch signalling, the availability of nuclear β-catenin is

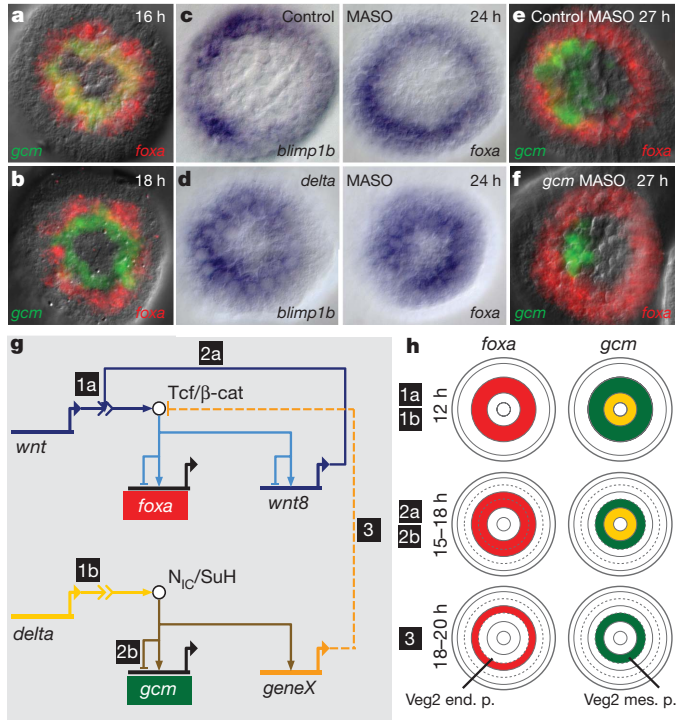


Figure 2 | Separation of endoderm and mesoderm regulatory states. **a, b**, DFISHs detecting expression of *foxa* and *gcm* at the indicated times post-fertilization. **c, d**, Demonstration that *delta* morpholino blocks mesodermal clearance of *foxa* and *blimp1b*; WMISH using *blimp1b* and *foxa* probes on control (**c**) and *delta* MASO-treated (**d**) embryos. **e, f**, DFISHs showing that the *foxa* expression pattern is independent of *Gcm* expression. **g**, Model of GRN interactions that determine the segregation of endodermal and mesodermal cell fates. Chronology is indicated by numbers 1–3. Wnt and Delta signals emanate from skeletogenic micromeres; geneX mediates interference with β-catenin activity. N_{IC}/SuH, Su(H)–intracellular Notch domain complex; Tcf/β-cat, Tcf/β-catenin complex. **h**, Schematic showing spatial patterns of gene expression in the developing embryo, with chronology as in **g**. Note the additional ring of veg2 descendants after 15 h, arising by radial cleavage; end. p., endoderm precursors; mes. p., mesoderm precursors.

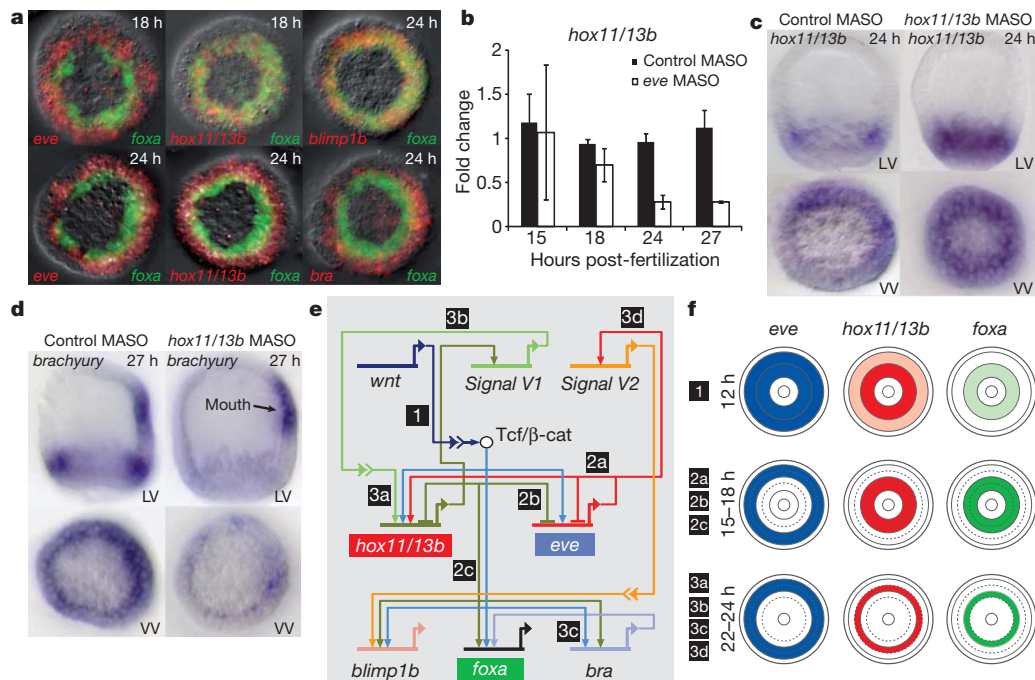


Figure 3 | Separation of anterior and posterior endoderm regulatory states. **a**, DFISHs showing the dynamic change in gene expression patterns. At 18 h, *hox11/13b* is co-expressed with *foxa* in veg2-derived cells, and *eve* is expressed in veg1 descendants exclusively with respect to *foxa*. At 24 h, *hox11/13b* and *brachyury* (*bra*) are expressed only in veg1-derived cells; *blimp1b* and *foxa* expression continue to overlap. **b**, Expression of *hox11/13b* in veg1 domain depends on *eve* expression. The histogram shows quantitative PCR

measurements of *hox11/13b* \pm s.d. ($n = 3$). **c**, Clearance of *hox11/13b* from veg2 descendants and its activation in veg1 descendants depend on Hox11/13b expression. LV, lateral view; VV, vegetal view. **d**, The expression of *brachyury* in veg1-derived cells depends on Hox11/13b expression. **e**, A chronological model of GRN interactions determining anterior versus posterior regulatory states. Chronology is indicated by numbers 1–3. **f**, Schematic showing spatial patterns of gene expression, with chronology as in **e**.

reduced, leading to Tcf/Groucho-mediated repression. This repression specifically affects veg2 endoderm regulatory genes. Thus, *hox11/13b* and *eve*, which are both expressed at this stage in veg1 endoderm progenitors, are not affected by interference with Delta/Notch signalling (Supplementary Fig. 6). The expression of the endoderm GRN in endoderm precursors is, in general, completely independent of Delta/Notch signalling (Supplementary Fig. 5).

In summary, the *cis*-regulatory Tcf responsiveness of early endodermal genes results first in the activation of an endodermal GRN and then, together with cleavage geometry, in the spatial separation of endodermal and mesodermal regulatory states and hence of biological fates (Fig. 2h). In contrast to many 'binary' cell-fate decisions that occur later²², this one involves no mutually acting repressors and no other bi-stable switch features. In fact, it is determinative, like much of early development, rather than bi-stable: there is no preceding intermediate state. Our results exclude an earlier model²³ proposing that clearance of *blimp1b* expression from the mesodermal domain^{19,24} is responsible for clearance of *wnt8* expression from this domain, on the assumption that Blimp1 is a necessary driver of *wnt8* expression. This could ultimately lead to the downregulation of most endodermal regulatory genes, by removal of the Tcf/ β -catenin signal that activates them. However, although mutation of Blimp1-binding sites reduces the activity of a small *wnt8 cis*-regulatory construct¹⁸, the same mutation does not affect expression of a bacterial artificial chromosome expression construct containing the whole genomic *wnt8 cis*-regulatory system (Supplementary Fig. 7). In any case, the expression of *wnt8* begins in veg2-derived cells long before the onset of *blimp1b* expression in these cells^{18,24}.

A few hours after the complete separation of endodermal and mesodermal cell fates, marked changes occur in the endodermal regulatory states. These result in the recruitment of two canonical hindgut regulatory genes into the veg1 endoderm GRN, which specifies future hindgut cell fate. By 24 h post-fertilization, *hox11/13b* and *brachyury*, which are both expressed in veg2 endoderm at 18 h, are being transcribed instead in veg1 endodermal progenitors, where *eve* also continues to

be expressed (Fig. 3a). The dynamic changes in the spatial expression of *hox11/13b* and *brachyury* can be explained by the results of perturbation experiments (Figs 1 and 3 and Supplementary Fig. 5). *Hox11/13b* expression is activated by Eve in descendants of veg1 cells at 24 h, because injection of *eve* MASO reduces *hox11/13b* expression only after 24 h (Fig. 3b). As seen previously for *blimp1b* and *eve*, auto-repression is required for the change in the *hox11/13b* expression domain and this auto-repression results in its clearance from veg2 endoderm. Accordingly, injection of *hox11/13b* MASO interferes with the clearance of *hox11/13b* transcripts from veg2 endoderm (Fig. 3c). In both the early veg2 and the later veg1 endoderm GRNs, Hox11/13b functions as a driver of *brachyury* expression, as shown by the specific reduction of endodermal *brachyury* expression in embryos injected with *hox11/13b* MASO (Fig. 3d).

Eve expression defines the veg1 regulatory state from 15 h post-fertilization but its detectable regulatory functions begin only after 24 h. The assembly of the veg1 endoderm GRN, which is spatially activated by Eve, is temporally motivated by a predicted signal (V1) expressed under the control of the veg2 endoderm GRN. We note that *hox11/13b* expression remains restricted to veg2 endoderm precursors in embryos injected with *hox11/13b* MASO (Fig. 3c). As summarized in Fig. 3e and f, the signal called V1, which is probably Wnt16 (data not shown), is expressed under the control of Hox11/13b in the veg2 lineage to induce *hox11/13b* expression in veg1 endoderm progenitors. There may be a signal from veg1 to veg2 as well: *blimp1b*, *brn1/2/4*, *gatae* and *tgif*, which continue to be expressed in veg2 endoderm, are indirectly affected by the knockdown of *eve* expression in veg1 descendants (Fig. 1f). A second putative signal (V2) is expressed under the control of Eve and activates expression of *blimp1b* and *gatae* in veg2 endoderm precursors. Blimp1 then activates *brn1/2/4* and *tgif* expression (Fig. 1f). As a possible consequence of signal V2, *blimp1b* expression becomes restricted to peripheral tiers of *foxa*-expressing cells just before gastrulation, when *blimp1b* transcripts accumulate in cells adjacent to the *eve* (and V2) expression domain (Supplementary

Fig. 3). *Blimp1b*, *gatae* and *tgif*, as well as the *tgif* driver gene *myc*, are all expressed in the midgut at the late gastrula stage and we propose that a future midgut regulatory state might be initiated here.

Figure 4 shows models of the ultimate anterior (veg2) and posterior (veg1) endoderm GRNs operating just before the onset of gastrulation. A few additional regulatory genes are activated in the final hours before gastrulation: *brn1/2/4*, *gatae*, *dac* and *tgif* are expressed in veg2-derived cells and *hnf1* is expressed in the veg1 endoderm domain by 27 h after fertilization. In the same time period, the number of direct cross-regulatory linkages increases markedly, as indicated by the results of perturbations shown in Fig. 1d–f (from 6 linkages at 18 h to 26 at 27 h). The models proposed here include previously identified linkages such as the positive feedback circuit between *blimp1b*, *otx* and *gatae*. Almost immediately, with the inception of gastrulation, the anterior endoderm GRN will be required to direct gastrular invagination and accordingly, this GRN achieves autonomy by this time point with the cross-regulatory structure shown in Fig. 4. In contrast, although the posterior endoderm network is uniquely specified, its structure is much simpler at this time because hindgut invagination is still many hours in the future.

Specific regulatory states thus distinguish anterior and posterior endoderm progenitors. These regulatory states are the outputs of GRNs composed of distinct sets of genes and regulatory interactions. Here we show, for the first time, the primary mechanistic basis for the different contributions of veg1 and veg2 endoderm to the future gut. Rather than the progressive differentiation of a broadly initiated, common ‘endomesoderm’ or ‘endoderm’ GRN, cell-fate specification results from the parallel activation of distinct GRNs, long before functional and morphological differentiation.

Regulatory system analysis generates a causal framework that extends vertically from the individual regulatory transactions encoded by the genome to the architecture of the control systems and thence to their ultimate outputs, the phenomena of dynamic spatial specification.

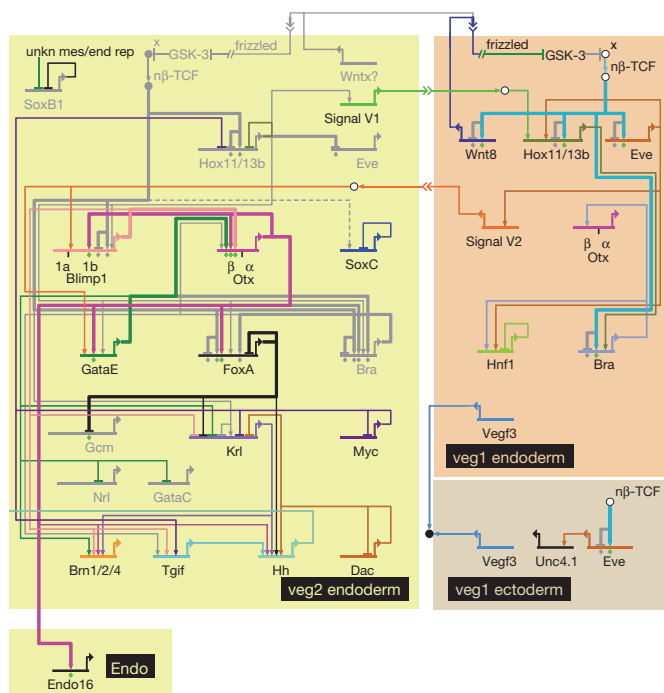


Figure 4 | Anterior and posterior GRNs just before gastrulation.

BioTapestry presentation as ‘view from the nucleus’ at 30 h (for an interactive version, including expression and perturbation results for each gene, as well as the temporal sequence of appearance or disappearance of linkages, see <http://supg.caltech.edu/endomes/#BioTapestryViewer>). Linkages shown in grey were active at earlier time points but by 30 h are extinguished. Unkn mes/end rep, unknown mesodermal and endodermal repressor of *soxB1*. For a discussion of the circuitry, see text.

This illuminates developmental biology in many ways. For example, we can now see why the endodermal cell lineages have different fates and how they acquire them; why and how the patterns of gene expression change; how the parts of the future gut are encoded and how they are pre-specified in a stepwise manner by the operation of the genomic regulatory system.

METHODS SUMMARY

All methods and procedures used in this work, including computational and data-reduction methods, have been published in detail earlier. Molecular biology procedures included digoxigenin-based single whole-mount *in situ* hybridization²⁵ and double-fluorescent¹⁷ whole-mount *in situ* hybridization, gene expression knockdown via validated morpholino oligonucleotides²⁶ and system-wide quantitative assessment of transcript levels in perturbed embryos²⁶. Computational procedures included the automated reduction of perturbation data (used to generate Supplementary Fig. 5) and the application of the GRN platform BioTapestry²⁷. Quantitative PCR probes and morpholino antisense oligonucleotides are listed in Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 December 2010; accepted 1 April 2011.

Published online 29 May 2011.

- Cameron, R. A., Fraser, S. E., Britten, R. J. & Davidson, E. H. Macromere cell fates during sea urchin development. *Development* **113**, 1085–1091 (1991).
- Logan, C. Y. & McClay, D. R. The allocation of early blastomeres to the ectoderm and endoderm is variable in the sea urchin embryo. *Development* **124**, 2213–2223 (1997).
- Ransick, A. & Davidson, E. H. Late specification of Veg1 lineages to endodermal fate in the sea urchin embryo. *Dev. Biol.* **195**, 38–48 (1998).
- Byrum, C. A., Xu, R., Bince, J. M., McClay, D. R. & Wikramanayake, A. H. Blocking Dishevelled signaling in the noncanonical Wnt pathway in sea urchins disrupts endoderm formation and spiculogenesis, but not secondary mesoderm formation. *Dev. Dyn.* **238**, 1649–1665 (2009).
- Logan, C. Y., Miller, J. R., Ferkowicz, M. J. & McClay, D. R. Nuclear β -catenin is required to specify vegetal cell fates in the sea urchin embryo. *Development* **126**, 345–357 (1999).
- Wikramanayake, A. H. *et al.* Nuclear β -catenin-dependent Wnt8 signaling in vegetal cells of the early sea urchin embryo regulates gastrulation and differentiation of endoderm and mesodermal cell lineages. *Genesis* **39**, 194–205 (2004).
- Sherwood, D. R. & McClay, D. R. LvNotch signaling mediates secondary mesenchyme specification in the sea urchin embryo. *Development* **126**, 1703–1713 (1999).
- Sweet, H. C., Hodor, P. G. & Ettensohn, C. A. The role of micromere signaling in Notch activation and mesoderm specification during sea urchin embryogenesis. *Development* **126**, 5255–5265 (1999).
- Sweet, H. C., Gehring, M. & Ettensohn, C. A. LvDelta is a mesoderm-inducing signal in the sea urchin embryo and can endow blastomeres with organizer-like properties. *Development* **129**, 1945–1955 (2002).
- Ransick, A. & Davidson, E. H. *cis*-regulatory processing of Notch signaling input to the sea urchin *glial cells missing* gene during mesoderm specification. *Dev. Biol.* **297**, 587–602 (2006).
- de-Leon, S. B. & Davidson, E. H. Information processing at the *foxa* node of the sea urchin endomesoderm specification network. *Proc. Natl Acad. Sci. USA* **107**, 10103–10108 (2010).
- Howard-Ashby, M. *et al.* Identification and characterization of homeobox transcription factor genes in *Strongylocentrotus purpuratus*, and their expression in embryonic development. *Dev. Biol.* **300**, 74–89 (2006).
- Howard-Ashby, M. *et al.* Gene families encoding transcription factors expressed in early development of *Strongylocentrotus purpuratus*. *Dev. Biol.* **300**, 90–107 (2006).
- Materna, S. C., Howard-Ashby, M., Gray, R. F. & Davidson, E. H. The C₂H₂ zinc finger genes of *Strongylocentrotus purpuratus* and their expression in embryonic development. *Dev. Biol.* **300**, 108–120 (2006).
- Rizzo, F., Fernandez-Serra, M., Squarzon, P., Archimandritis, A. & Arnone, M. I. Identification and developmental expression of the ets gene family in the sea urchin (*Strongylocentrotus purpuratus*). *Dev. Biol.* **300**, 35–48 (2006).
- Tu, Q., Brown, C. T., Davidson, E. H. & Oliveri, P. Sea urchin Forkhead gene family: phylogeny and embryonic expression. *Dev. Biol.* **300**, 49–62 (2006).
- Peter, I. S. & Davidson, E. H. The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage. *Dev. Biol.* **340**, 188–199 (2010).
- Minokawa, T., Wikramanayake, A. H. & Davidson, E. H. *cis*-Regulatory inputs of the *wnt8* gene in the sea urchin endomesoderm network. *Dev. Biol.* **288**, 545–558 (2005).
- Smith, J., Kraemer, E., Liu, H., Theodoris, C. & Davidson, E. A spatially dynamic cohort of regulatory genes in the endomesodermal gene network of the sea urchin embryo. *Dev. Biol.* **313**, 863–875 (2008).
- Peter, I. S. & Davidson, E. H. Modularity and design principles in the sea urchin embryo gene regulatory network. *FEBS Lett.* **583**, 3948–3958 (2009).
- Croce, J. C. & McClay, D. R. Dynamics of Delta/Notch signaling on endomesoderm segregation in the sea urchin embryo. *Development* **137**, 83–91 (2010).

22. Davidson, E. H. Emerging properties of animal gene regulatory networks. *Nature* **468**, 911–920 (2010).
23. Smith, J., Theodoris, C. & Davidson, E. H. A gene regulatory network subcircuit drives a dynamic pattern of gene expression. *Science* **318**, 794–797 (2007).
24. Livi, C. B. & Davidson, E. H. Regulation of *spblimp1/krox1a*, an alternatively transcribed isoform expressed in midgut and hindgut of the sea urchin gastrula. *Gene Expr. Patterns* **7**, 1–7 (2007).
25. Ransick, A. Detection of mRNA by *in situ* hybridization and RT-PCR. *Methods Cell Biol.* **74**, 601–620 (2004).
26. Oliveri, P., Tu, Q. & Davidson, E. H. Global regulatory logic for specification of an embryonic cell lineage. *Proc. Natl Acad. Sci. USA* **105**, 5955–5962 (2008).
27. Longabaugh, W. J., Davidson, E. H. & Bolouri, H. Visualization, documentation, analysis, and communication of large-scale gene regulatory networks. *Biochim. Biophys. Acta* **1789**, 363–374 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge technical assistance from J. Yun, who executed much of the perturbation analysis matrix, and from A. Puszyńska and E. Erkenbrack, who contributed to whole-mount *in situ* hybridization results. We are grateful to E. Rothenberg for a detailed critique of the manuscript. I.S.P. was the recipient of a fellowship from the Swiss National Science Foundation in the initial stages of this work. The research was supported by National Institutes of Health grant HD37105 to E.H.D.

Author Contributions This research was conceived by I.S.P. and E.H.D. and all experiments were designed and executed by I.S.P. with the assistance acknowledged above. The results were interpreted by I.S.P. and E.H.D., who also contributed jointly to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to E.H.D. (davidson@caltech.edu) or I.S.P. (ipeter@caltech.edu).

METHODS

MASO injection and RNA isolation. MASOs were provided by GeneTools and sequences are given in Supplementary Table 5. MASOs were microinjected into fertilized eggs at 100–400 μM in 0.12 M KCl, as described in ref. 26 and 28. Control MASOs consisting of random sequences 25 nucleotides long were injected at the same or higher concentrations as gene-specific MASOs. Experiments were performed on 2–4 independent embryonic batches. Embryos were cultured at 15 °C and about 100–200 embryos per sample were lysed at 24 h or 27 h after fertilization. Total RNA was isolated using the RNeasy Micro Kit (Qiagen).

Quantitative PCR analysis. Complementary DNA was synthesized using the iScript cDNA synthesis kit (BioRad). Quantitative PCR was performed using iTaq SYBR green supermix (BioRad) on an amount of cDNA equivalent to 0.6 embryos in a 10 μl reaction with the gene-specific primers listed in Supplementary Table 4. Gene expression levels were normalized to levels of ubiquitin expression. Changes in expression levels were determined by comparing normalized expression levels in MASO-injected embryos to normalized expression levels in un-injected control embryos²⁹. Changes were considered significant if gene expression levels decreased more than threefold or increased more than twofold in embryos injected with a gene-specific MASO. Target gene expression was considered to be affected by a regulatory factor at a given time point if the injection of a regulatory-gene-specific MASO, but not the injection of control MASOs, resulted in significant changes in target gene expression in the majority of experiments. Computational procedures included the automated reduction of perturbation data (used to generate Supplementary Fig. 5) and the application of the GRN platform BioTapestry²⁷.

Whole-mount *in situ* hybridization. Probe templates were either derived from a cDNA library or generated by PCR amplification of cDNA synthesized from the RNA of 27 h embryos with gene-specific primers listed in Supplementary Table 6. Antisense RNA probes labelled with digoxigenin or fluorescein were generated using the corresponding RNA labelling mix (Roche). Whole-mount *in situ* hybridizations (ISHs) were performed according to standard methods^{25,30}. Briefly, embryos were fixed in 4% paraformaldehyde, 32.5% sea water, 32.5 mM maleic acid (pH 7) and 162.5 mM NaCl at 4 °C overnight. Hybridizations were performed overnight at 65 °C using a probe concentration of 1 ng μl^{-1} . For single whole-mount ISH, probes were detected using anti-digoxigenin Fab fragments conjugated to alkaline phosphatase (1:1,000 dilution) and NBT/BCIP (nitro-blue tetrazolium chloride/ 5-bromo-4-chloro-3'-indolylphosphate p-toluidine salt). Hybridizations for double whole-mount ISH were performed according to a protocol described previously¹⁸. Probes were detected by horseradish-peroxidase-conjugated anti-digoxigenin or anti-fluorescein Fab fragments (1:1,000 dilution) using substrates provided in the TSA Plus Cyanine3/Fluorescein System (Perkin Elmer). Staining occurred at a substrate dilution of 1:400 in 1× Plus Amplification Diluent or in Tris-buffered saline with 0.005% H_2O_2 .

28. Rast, J. P. *et al.* Recovery of developmentally defined gene sets from high-density cDNA macroarrays. *Dev. Biol.* **228**, 270–286 (2000).
29. Materna, S. C. & Oliveri, P. A protocol for unraveling gene regulatory networks. *Nature Protocols* **3**, 1876–1887 (2008).
30. Revilla-i-Domingo, R., Oliveri, P. & Davidson, E. H. A missing link in the sea urchin embryo gene regulatory network: *hesC* and the double-negative specification of micromeres. *Proc. Natl Acad. Sci. USA* **104**, 12383–12388 (2007).

Condensin association with histone H2A shapes mitotic chromosomes

Kenji Tada^{1,2}, Hiroaki Susumu^{1,2}, Takeshi Sakuno^{1,3} & Yoshinori Watanabe^{1,2}

Chromosome structure is dynamically regulated during cell division, and this regulation is dependent, in part, on condensin. The localization of condensin at chromosome arms is crucial for chromosome partitioning during anaphase. Condensin is also enriched at kinetochores but its precise role and loading machinery remain unclear. Here we show that fission yeast (*Schizosaccharomyces pombe*) kinetochore proteins Pcs1 and Mde4—homologues of budding yeast (*Saccharomyces cerevisiae*) monopolin subunits and known to prevent merotelic kinetochore orientation—act as a condensin ‘recruiter’ at kinetochores, and that condensin itself may act to clamp microtubule binding sites during metaphase. In addition to the regional recruitment factors, overall condensin association with chromatin is governed by the chromosomal passenger kinase Aurora B. Aurora-B-dependent phosphorylation of condensin promotes its association with histone H2A and H2A.Z, which we identify as conserved chromatin ‘receptors’ of condensin. Condensin phosphorylation and its deposition onto chromosome arms reach a peak during anaphase, when Aurora B kinase relocates from centromeres to the spindle midzone, where the separating chromosome arms are positioned. Our results elucidate the molecular basis for the spatiotemporal regulation of mitotic chromosome architecture, which is crucial for chromosome partitioning.

Chromosomes carry genetic information and change their shape during the cell cycle. During mitosis, chromosomes are packaged into rod-shaped structures because the interphase (transcriptionally active) chromosomes are too diffuse to be separated along the spindle at anaphase. Condensin, a major protein component of mitotic chromosomes, is required for the assembly of rigid chromosomes and their separation^{1–3}. Condensation of centromeric chromatin may also be important because depletion of condensin impairs the stiffness of centromeres and causes dysfunction of kinetochores^{4–8}. However, this explanation was obtained by depleting whole cellular condensin rather than the centromere-specific pool, leaving the precise role of condensin at centromeres elusive.

Unlike cohesin, which embraces DNA topologically to establish sister-chromatid cohesion⁹, condensin may regulate chromosome architecture through a dynamic (non-topological) interaction with chromatin^{10,11}. Studies in budding yeast and fission yeast identified the transcription factor of polymerase III (Pol III) as a potential condensin loader^{12,13}, whereas in budding yeast condensin loading in rDNA regions depends on monopolin¹⁴, a factor also required for mono-orientation of sister chromatids at meiosis I (ref. 15). In vertebrates, the non-catalytic action of PP2A and the methylation of histone H4 lysine 20 are implicated in the chromosomal association of non-canonical condensin (condensin II)^{16,17}, whereas these regulations are not applicable to canonical condensin (condensin I), which is conserved in all eukaryotic organisms.

Since the coincidence of histone phosphorylation and chromosome condensation during the cell cycle was recognized, it has been suspected that these two events are linked^{1,18}. The conserved mitotic kinase Aurora B, which also phosphorylates histone H3S10, regulates several crucial mitotic events, including chromosome orientation and cytokinesis¹⁹. Aurora B is important for chromosome condensation or the association of condensin with chromatin in a wide range of species (including vertebrate condensin I)^{8,20–27}, whereas H3S10 phosphorylation might

not be responsible for this function^{28,29}. Thus, because the key target of Aurora B kinase in chromosome condensation has not been identified, the link between Aurora B and chromosome architecture remains a crucial issue^{1,2,18}.

Condensin may clamp microtubule binding sites

To address mechanisms of condensin loading to chromatin, we first dissected the functions of condensin at kinetochores and chromosome arm regions in fission yeast. By inactivating only the kinetochore pool of condensin in fission yeast, we demonstrate that condensin has a specific role at kinetochores in preventing merotelic attachment, most probably by assembling rigid kinetochores (Supplementary Fig. 1). The fission yeast kinetochore proteins Pcs1 and Mde4, which are homologues of budding yeast monopolin subunits Csm1 and Lrs4 but are dispensable for mono-orientation at meiosis I, are reportedly required for preventing merotelic attachment^{15,30,31}. Thus, the Pcs1–Mde4 complex and condensin, both enriched at the kinetochore^{8,31}, might cooperate in a similar function. Supporting this assumption, we detected a physical interaction between the Pcs1–Mde4 complex and condensin (Supplementary Fig. 2) as well as synthetic growth defects in *pcs1Δ* and the condensin mutation *cut3-477* (Fig. 1a). To examine the localization interdependency, we performed a chromatin immunoprecipitation (ChIP) assay using prometaphase-arrested cells. Condensin is enriched at the kinetochore and rDNA region, and moderately along chromosome arms, as reported previously^{8,12,13}. Notably, kinetochore condensin localization, but not pericentric localization, is diminished in *pcs1Δ* cells (Fig. 1b and Supplementary Fig. 3), whereas monopolin localization is intact in *cut3-477* cells (Supplementary Fig. 4a). Although condensin at the rDNA region is also diminished in *pcs1Δ* cells (Fig. 1b), no selective defects in anaphase separation of the rDNA region were observed in *pcs1Δ* or *cut3-477* cells (Supplementary Fig. 5). We confirmed that Aurora B localization, as well as its activity at centromeres, is intact in *pcs1Δ* cells (Supplementary Fig. 6), precluding

¹Laboratory of Chromosome Dynamics, Institute of Molecular and Cellular Biosciences, University of Tokyo, Yayoi, Tokyo 113-0032, Japan. ²Graduate Program in Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Yayoi, Tokyo 113-0032, Japan. ³Promotion of Independence for Young Investigators, University of Tokyo, Yayoi, Tokyo 113-0032, Japan.

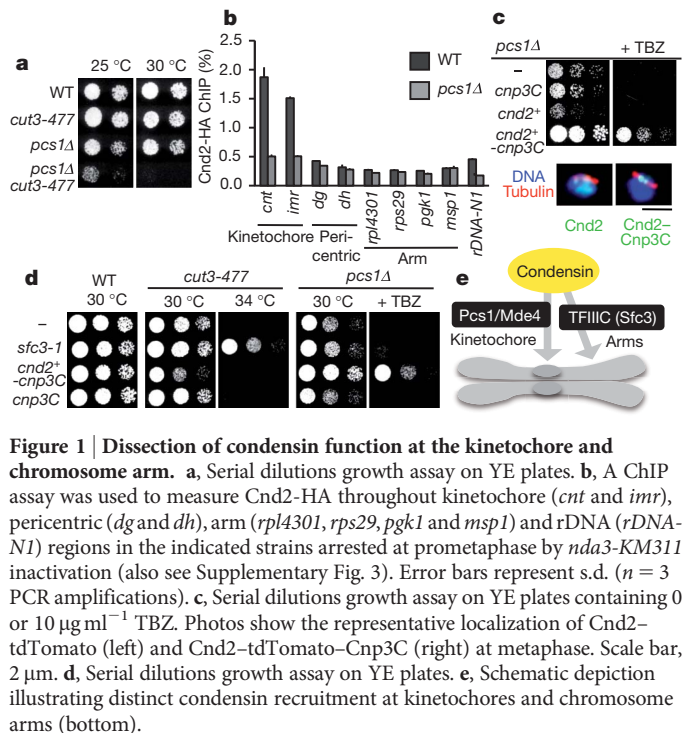


Figure 1 | Dissection of condensin function at the kinetochore and chromosome arm. **a**, Serial dilutions growth assay on YE plates. **b**, A ChIP assay was used to measure Cnd2-HA throughout kinetochore (*cnt* and *imr*), pericentric (*dg* and *dh*), arm (*rpl4301*, *rps29*, *pgk1* and *msp1*) and rDNA (*rDNA-N1*) regions in the indicated strains arrested at prometaphase by *nda3-KM311* inactivation (also see Supplementary Fig. 3). Error bars represent s.d. ($n = 3$ PCR amplifications). **c**, Serial dilutions growth assay on YE plates containing 0 or 10 $\mu\text{g ml}^{-1}$ TBZ. Photos show the representative localization of Cnd2-tdTomato (left) and Cnd2-tdTomato-Cnp3C (right) at metaphase. Scale bar, 2 μm . **d**, Serial dilutions growth assay on YE plates. **e**, Schematic depiction illustrating distinct condensin recruitment at kinetochores and chromosome arms (bottom).

the possibility that merotelic attachment is caused by Aurora B dysfunction at centromeres.

To validate further the importance of condensin at kinetochores in the context of Pcs1-Mde4, we engineered a Cnd2-Cnp3C fusion protein, which targets kinetochores through the Cnp3C domain³², even in *pcs1Δ* cells. The expression of this fusion protein largely

suppresses the growth defect, sensitivity to thiabendazole (TBZ, a microtubule destabilizing drug) and the incidence of lagging chromosomes in *pcs1Δ* cells, whereas neither Cnd2 nor Cnp3C protein alone suppresses these phenotypes (Fig. 1c and Supplementary Fig. 7). These results suggest that condensin itself may have a role in clamping microtubule attachment sites to prevent merotelic attachment, the function previously attributed to monopolin^{30,31}.

Dissection of condensin function

The condensin mutant *cut3-477*, which impairs, at least, condensin localization along whole chromosome regions (Supplementary Fig. 4a), shows a defect mainly in anaphase separation of the chromosome arms³³ (Supplementary Fig. 5). Indeed, the growth defect in *cut3-477* cells is not suppressed by centromere tethering of condensin (Fig. 1d), indicating that condensin at the kinetochore is not rate-limiting for the viability of *cut3-477* cells. Recent studies suggest that condensin loading along the chromosome arms depends on the Pol III transcription factor TFIIC, the targets of which are distributed along the chromosome arms including pericentric tRNA clusters^{12,13,34}. As reported previously¹², the *cut3-477* mutant is suppressed by *sfc3-1*, a 'gain-of-function' mutation of TFIIC, by which condensin localization is elevated along chromosome arms but not at the kinetochores and rDNA regions (Fig. 1d and Supplementary Fig. 8). In contrast, the growth defect and TBZ sensitivity of *pcs1Δ* cells are not suppressed by the *sfc3-1* mutation, whereas they are suppressed by an increase in kinetochore condensin (Fig. 1d). Taken together, these results indicate that the pools of condensin at kinetochores and along the chromosome arms, which are recruited respectively by the Pcs1-Mde4 complex and TFIIC, are functionally separable (Fig. 1e).

Aurora-B-dependent condensin localization

Given that the mitotic kinase Aurora B may regulate condensin localization to chromatin in fission yeast^{8,23}, we re-investigated condensin

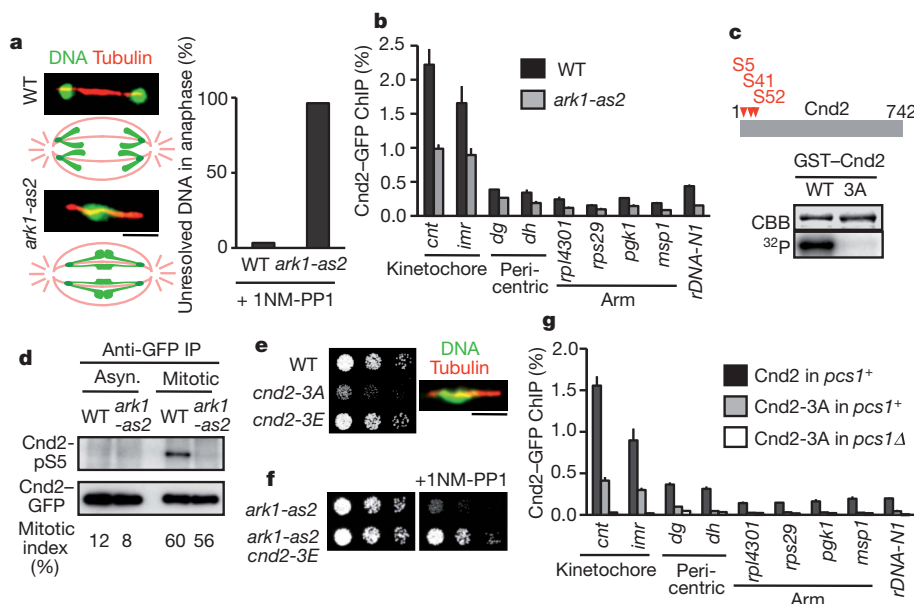


Figure 2 | Phosphorylation of condensin by Aurora B is required for chromatin association. **a**, Frequencies of unresolved DNA at anaphase ($n > 100$ cells) were examined in the indicated cells. **b**, Wild-type (WT) and *ark1-as2* cells expressing Cnd2-GFP were arrested at metaphase by the overexpression of *mad2+*. For the last 1 h, cells were cultured in the presence of 1NM-PP1, then fixed and examined by ChIP assay. Error bars represent s.d. ($n = 3$ PCR amplifications). **c**, Schematic diagram of fission yeast Cnd2. Aurora B phosphorylation sites are indicated by arrowheads. GST-Cnd2 and GST-Cnd2-3A (alanine substitution at S5, S41 and S52) were phosphorylated by Ark1 and analysed for phosphate incorporation (³²P) and protein levels (Coomassie brilliant blue; CBB). **d**, Cnd2-GFP proteins were precipitated from

asynchronous and mitotic (arrested by the overexpression of *mad2+*) extracts of wild-type and *ark1-as2* cells, and analysed by immunoblot using anti-Cnd2-pS5 or anti-GFP antibodies. For the last 30 min of mitotic arrest, cells were cultured in the presence of 5 μM 1NM-PP1. The mitotic index was determined by counting nuclear Cnd2-GFP-positive cells. **e**, Serial dilution growth assay on YE plates. Photo shows a representative *cnd2-3A* cell at anaphase. **f**, Serial dilution growth assay on YE plates in the presence or absence of 0.2 μM 1NM-PP1. **g**, The indicated cells ectopically expressing Cnd2-GFP or Cnd2-3A-GFP were arrested at prometaphase by *nda3-KM311* inactivation and examined by ChIP assay. Error bars represent s.d. ($n = 3$ PCR amplifications). Scale bars, 2 μm (**a**, **e**).

loading using the ATP-analogue (1NM-PP1)-sensitive Aurora B mutant *ark1-as2* (ref. 35), which showed an extensive defect in chromosome separation at anaphase only when treated with 1NM-PP1 (Fig. 2a). A ChIP assay indicated that condensin localization decreases in *ark1-as2* cells in both the kinetochore and arm regions, whereas Mde4 localization is preserved intact (Fig. 2b and Supplementary Fig. 4b). Thus, Aurora B might act on condensin itself rather than the region-specific recruiter, Pcs1–Mde4 or TFIIC. By performing *in vitro* phosphorylation of all the condensin subunits purified from bacteria, we identified the amino-terminal fragment of kleisin subunit Cnd2 (Cnd2(N)), which carries several Aurora B consensus sites³⁶ (Supplementary Fig. 9a). *In vitro* phosphorylation assays using mutant Cnd2 proteins indicate that S5, S41 and S52 are the major Aurora B phosphorylation sites (Fig. 2c and Supplementary Fig. 9b). A phospho-specific antibody raised against one of these sites, Cnd2-pS5 (Supplementary Fig. 10), recognized Cnd2 prepared from mitotic cell extract, and this phosphorylation was abolished by inactivation of Ark1 (Fig. 2d). We conclude that Cnd2 is phosphorylated depending on the mitotic kinase Aurora B *in vivo*.

To examine the functional significance of Cnd2 phosphorylation, we replaced the endogenous *cnd2*⁺ with the non-phosphorylatable mutant *cnd2-3A*. The *cnd2-3A* mutant showed a severe growth defect, unresolved DNA signal during anaphase and a high incidence of lagging chromosomes, hallmarks of condensin defects (Fig. 2e and Supplementary Fig. 11). A ChIP assay indicates that the chromatin

localization of Cnd2-3A is largely reduced at all tested chromosome loci (Fig. 2g). This defect is not due to the inability of the Cnd2-3A protein to form the condensin complex or to localize to the nucleus (Supplementary Fig. 12). Notably, the phospho-mimetic mutant *cnd2-3E* suppresses the growth defect of *ark1-as2* cells (Fig. 2f), indicating that the phosphorylation of condensin is the important essential target of Aurora B in fission yeast.

In human cells, chromatin association of condensin I is regulated by Aurora B (refs 25, 26) (Supplementary Fig. 13). Indeed, bacterially purified CAP-H (kleisin of condensin I), but not CAP-H2 (kleisin of condensin II), is phosphorylated by Aurora B *in vitro*; this phosphorylation is largely attributable to the serine 70 residue of CAP-H (Fig. 3a and Supplementary Fig. 14). To examine *in vivo* phosphorylation, we raised a phospho-specific antibody against CAP-H-pS70 (Supplementary Fig. 15). Immunoblot assays revealed that the phosphorylation of CAP-H at serine 70 is mitosis-specific and is Aurora-B-dependent (Fig. 3b), and that phosphorylated CAP-H is selectively enriched in the chromatin of mitotic cells (Fig. 3c). To examine the significance of Aurora-B-dependent phosphorylation of CAP-H for chromatin binding, we expressed an EGFP-tagged non-phosphorylatable CAP-H-S70A protein and compared it with its wild-type counterpart in HeLa cells (Fig. 3d). Whereas both the wild-type and mutant CAP-H proteins could form intact condensin complexes *in vivo* (Fig. 3e), the CAP-H-S70A protein showed a defect in chromatin localization. Treatment of mitotic cells with the Aurora B inhibitor caused a similar

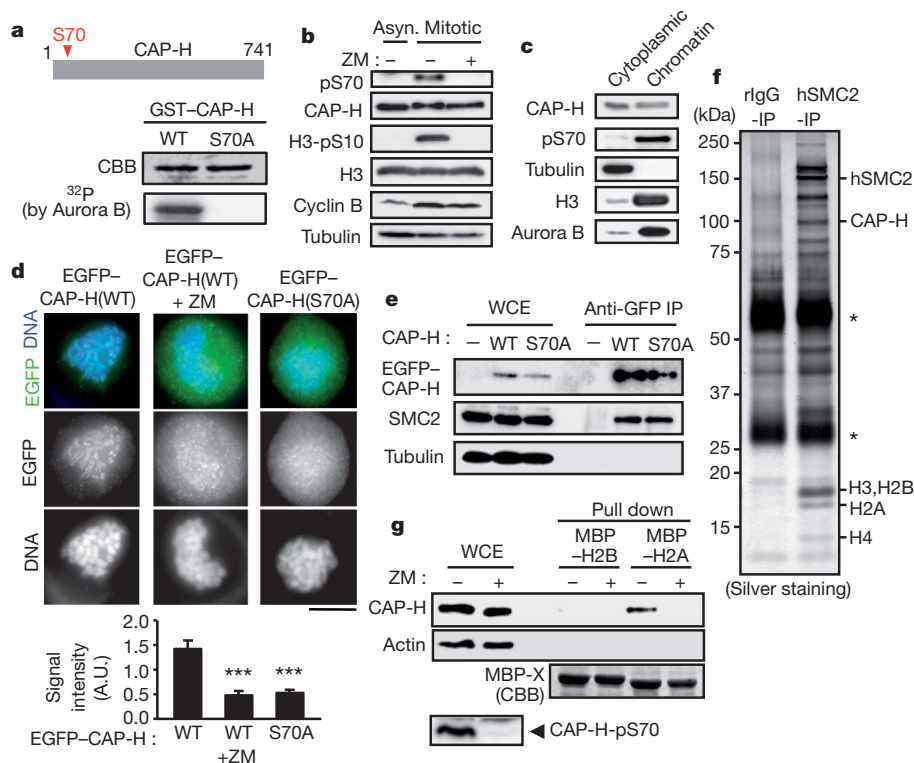


Figure 3 | Chromatin association of human condensin I is also regulated by Aurora-B-dependent phosphorylation. **a**, GST–CAP-H wild-type and GST–CAP-H-S70A were incubated with GST–Aurora B, and analysed for phosphate incorporation (³²P) and protein levels (CBB). **b**, Whole-cell extracts were obtained from HeLa cells treated with or without the Aurora inhibitor ZM447439 (ZM) at prometaphase or asynchronous cells, and analysed by immunoblot using the indicated antibodies. **c**, Cell extracts prepared from nocodazole-arrested cells were fractionated into chromatin and cytoplasmic fractions, and analysed by immunoblot. **d**, Synchronized mitotic HeLa cells expressing GFP–CAP-H were treated with nocodazole in the absence or presence of ZM447439 for 2 h. Signals of GFP–CAP-H at prometaphase were examined by immunostaining. Cells expressing GFP–CAP-H-S70A were similarly analysed without ZM447439

treatment. The ratios of nucleus and cytoplasmic signals were quantified in the indicated cells. Error bars represent s.e.m. ($n = 15$ cells). ***, $P < 0.001$. Scale bar, 5 μ m. **e**, Co-immunoprecipitation of GFP–CAP-H and SMC2. Whole-cell extracts (WCE) were prepared from nocodazole-arrested HeLa cells expressing GFP–CAP-H wild type or GFP–CAP-H-S70A. GFP–CAP-H was immunoprecipitated (IP) with anti-GFP antibodies to examine the co-precipitation of SMC2. **f**, Chromatin fractions were prepared from nocodazole-arrested cells, precipitated with control or anti-SMC2 antibodies, and analysed by silver staining. Asterisks indicate IgG. **g**, Whole-cell extracts were prepared from mitotic HeLa cells treated with or without ZM447439, pulled down with MBP (maltose-binding protein)–H2A or MBP–H2B, and analysed by immunoblot using antibodies against CAP-H, CAP-H-pS70 or actin.

defect even in the wild-type CAP-H protein (Fig. 3d). We conclude that the association of condensin I with chromatin largely relies on the Aurora-B-dependent phosphorylation of the kleisin subunit in human cells, as in fission yeast.

Condensin associates with histone H2A

The foregoing results indicate that the Aurora-B-dependent phosphorylation of Cnd2/CAP-H might promote its interaction with one or more chromatin factors that locate commonly at the kinetochores and arm regions. To identify such factors, we immunoprecipitated condensin from mitotic HeLa cell extracts using anti-Smc2 antibodies. Silver staining of the immunoprecipitates detected most condensin subunits and several additional bands in the 10–20 kDa range, which correspond to histones (Fig. 3f and Supplementary Fig. 16). To test the possibility that a histone acts as the chromatin receptor of CAP-H, we performed a pull-down assay using recombinant CAP-H and histones. Remarkably, CAP-H, but not CAP-H2, interacted specifically with histone H2A and its variant H2A.Z *in vitro* (Supplementary Fig. 17a). Furthermore, histone H2A pulled down CAP-H from mitotic

cell extracts, but failed to do so when cell extracts were prepared after the inhibition of Aurora B (Fig. 3g), indicating that CAP-H associates with H2A in an Aurora-B-dependent manner. Although condensin may associate with DNA^{1,2}, the detected interaction between H2A and condensin is not mediated by DNA (Supplementary Fig. 18). Although the depletion of H2A from cells is technically difficult, we were able to deplete H2A.Z by means of an RNA interference (RNAi) method reported previously³⁷. Consequently, condensin localization was significantly decreased in H2A.Z RNAi prometaphase cells, although a substantial amount of condensin still localized on the chromatin, presumably because of intact H2A (Supplementary Fig. 17b). This result fits with the recent finding that the depletion of H2A.Z partially impairs condensin localization in fission yeast³⁸.

To delineate the significance of the interaction between the condensin kleisin subunit and H2A or H2A.Z, we performed comparable analyses in fission yeast, a genetically tractable organism. Fission yeast Cnd2 interacted with H2A and H2A.Z *in vitro*, and these interactions were enhanced when Cnd2 was phosphorylated by Aurora B (Fig. 4a, b and Supplementary Fig. 19a). The interaction with Cnd2 is mediated by

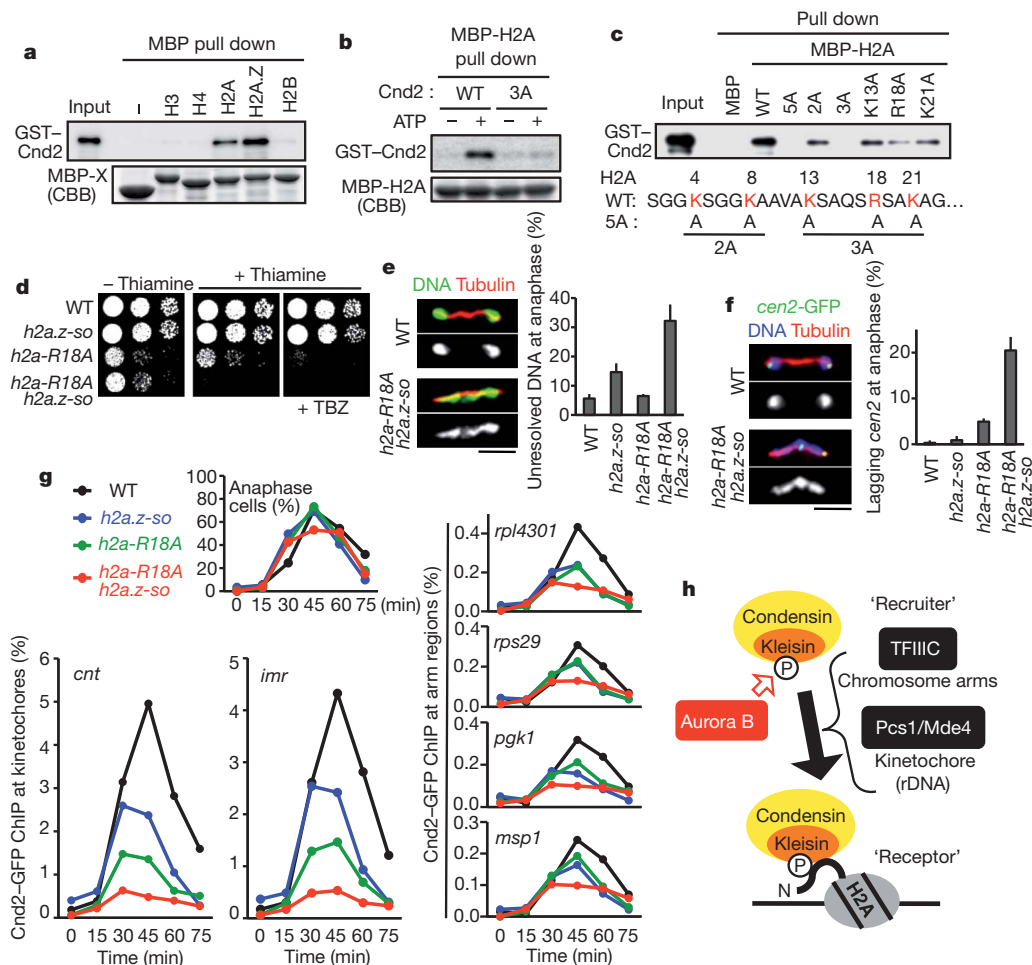


Figure 4 | H2A and H2A.Z act as a chromatin receptor of condensin.

a, GST-Cnd2 phosphorylated by Ark1 was pulled down with the indicated MBP-fused proteins and analysed by immunoblot using anti-GST antibody. **b**, GST-Cnd2 wild type or GST-Cnd2-3A reacted with Ark1 in the absence or presence of ATP was pulled down with MBP-H2A. **c**, GST-Cnd2 phosphorylated by Ark1 was pulled down with the indicated MBP-fused proteins. Mutated residues within the N-terminal tail of H2A are shown in the diagram. **d**, Serial dilution growth assay on SSA plates containing 0 or 8 μ M thiamine and 0 or 10 μ g ml⁻¹ TBZ. **e**, The indicated cells were cultured in SSA medium (–thiamine), transferred to YE medium (+thiamine) and incubated for 36 h at 30 °C to suppress the expression of H2A.Z. Frequencies of

unresolved DNA at anaphase ($n > 100$ cells) were examined. Error bars represent s.d. ($n = 3$ experiments). **f**, The indicated cells cultured in YE medium for 24 h at 30 °C were shifted to 18 °C for 6 h. Frequencies of lagging chromosomes of cen2-GFP at anaphase ($n > 100$ cells) were examined. Error bars represent s.d. ($n = 3$ experiments). **g**, The indicated cells were arrested at G2 phase by *cdc25-22* inactivation and released into mitosis. Cells collected at each time point were examined by ChIP assay (average of two PCR amplifications). Anaphase cells were monitored by DAPI staining ($n > 100$ cells). **h**, Schematic depiction of the pathway that regulates condensin localization on chromatin. Scale bars, 2 μ m (**e**, **f**).

the N-terminal tail of H2A or H2A.Z, which is enriched in basic amino acid residues (Fig. 4c and Supplementary Fig. 19b). Accordingly, an H2A-3A mutant protein in which three lysine/arginine residues (K13, R18 and K21) were replaced with alanines lost the affinity for phosphorylated Cnd2 (Fig. 4c). Among these three residues, arginine 18 seemed to be most important for the interaction with Cnd2.

H2A and H2A.Z act as condensin receptors

To explore the significance of the binding between Cnd2 and H2A *in vivo*, we tried to replace two fission yeast H2A genes (*hta1*⁺ and *hta2*⁺) with mutant alleles. The H2A-3A protein, which entirely abolishes the *in vitro* binding with Cnd2, causes cell lethality (not shown). Instead, we used mutant cells in which both histone H2A genes were replaced with the H2A-R18A mutant allele (called *h2a-R18A*, hereafter). The *h2a-R18A* cells showed a growth defect as well as sensitivity to TBZ (Fig. 4d). In contrast, depletion of the H2A.Z gene (*pht1*⁺) resulted in no TBZ sensitivity, whereas this mutation showed synthetic lethality with the *h2a-R18A* mutation (not shown). To explore the synthetic defects, we replaced the *pht1*⁺ promoter with a conditionally repressible promoter to shut off (so) the expression of H2A.Z (*h2a.z-so*), and crossed it with *h2a-R18A*. The double mutant showed lethality only under the H2A.Z shut-off condition (Fig. 4d), and exhibited unresolved DNA as well as lagging of *cen2*-GFP signals at anaphase, a typical phenotype of the condensin mutant (Fig. 4e, f). Similar, if not identical, defects were observed when *h2a.z-so* was combined with *h2a-R18Q*, another mutation erasing the positive charge at R18, but not with *h2a-R18K* (Supplementary Fig. 20). Taken together, these results indicate that H2A and H2A.Z act redundantly in condensin function by directly interacting with condensin.

To explore condensin loading in the histone mutants, we arrested cells at the G2/M boundary, and then monitored condensin occupancy kinetics at the kinetochore and chromosome arm regions through synchronous mitosis. In wild-type cells condensin loading took place through prophase to metaphase and culminated in anaphase (Fig. 4g). In *h2a-R18A* cells, however, condensin loading was significantly decreased in all tested regions throughout mitosis (Fig. 4g). A similar defect was observed in *h2a.z-so* cells, which showed the condensin loading defect particularly in anaphase, as previously reported in H2A.Z-depleted cells³⁸. The decrease in condensin localization was much more marked in the *h2a-R18A h2a.z-so* double mutant than in either single mutant alone (Fig. 4g), indicating the redundancy of histone H2A and H2A.Z function in localizing condensin to chromatin. Notably, *h2a-R18A* cells reduced condensin and its function more at the kinetochores than the arm regions as compared with *h2a.z-so* cells (Fig. 4e–g). This can be explained by the fact that H2A.Z is less abundant around the centromeres³⁹ and so the contribution of H2A might be relatively larger in this region. We confirmed that the condensin loading defect in the *h2a-R18A* and *h2a.z-so* mutants is not due to decreased expression or defective complex formation of condensin (Supplementary Fig. 21). In contrast to condensin localization, the localization of the condensin recruiter Mde4 or cohesin is intact in *h2a-R18A* cells, indicative of the specificity (Supplementary Fig. 22). Taken together, these results support the notion that H2A and H2A.Z function as the chromatin receptor of condensin (Fig. 4h).

Spatiotemporal localization of condensin

The overall chromatin localization of condensin, which peaks in anaphase rather than metaphase (Fig. 4g), may reflect Aurora B activity towards condensin. Accordingly, phosphorylation of condensin by Aurora B culminates in anaphase (Fig. 5a). As a chromosomal passenger component, Aurora B localizes to centromeres in metaphase, and subsequently transfers to the spindle midzone in anaphase^{19,23}. Because the redistributed Aurora B forms a special phosphorylation gradient centred at the spindle midzone⁴⁰, this gradient might efficiently phosphorylate chromatin-free condensin, especially around the separating

chromosome arms. Indeed, if condensin association with chromatin is compared between metaphase and anaphase, the increase during anaphase is more pronounced at the arm regions than at the kinetochores (Supplementary Fig. 23).

The spatiotemporal condensin localization model implies that if chromosome separation is impeded along the chromosome arms at anaphase, condensin accumulation will be increased especially at the arm regions. To test this prediction, we used cells depleted for Taz1, a telomere binding protein, because at lower temperature this mutant shows entanglement at telomeres and interferes with chromosome separation in anaphase⁴¹. Indeed, *taz1Δ* cells show a transient anaphase bridging of chromosomes, although most chromosomes finally managed to separate presumably by pulling off the entanglement (Fig. 5b). Remarkably, condensin localization at anaphase increases along the chromosome arms in *taz1Δ* cells compared with wild-type cells, whereas the condensin profile at the kinetochores is invariable (Fig. 5b). Because accumulation of condensin I during anaphase is also reported in human cells^{10,26}, we compared the distribution of chromosomes and condensin I in mitotic HeLa cells. Immunostaining of early anaphase cells revealed that condensin I deposition onto chromatin is enhanced near the Aurora B sites (Supplementary Fig. 24). These results support the notion that Aurora B promotes chromosome separation near the spindle midzone by efficiently localizing condensin to the chromosome arms (Fig. 5c).

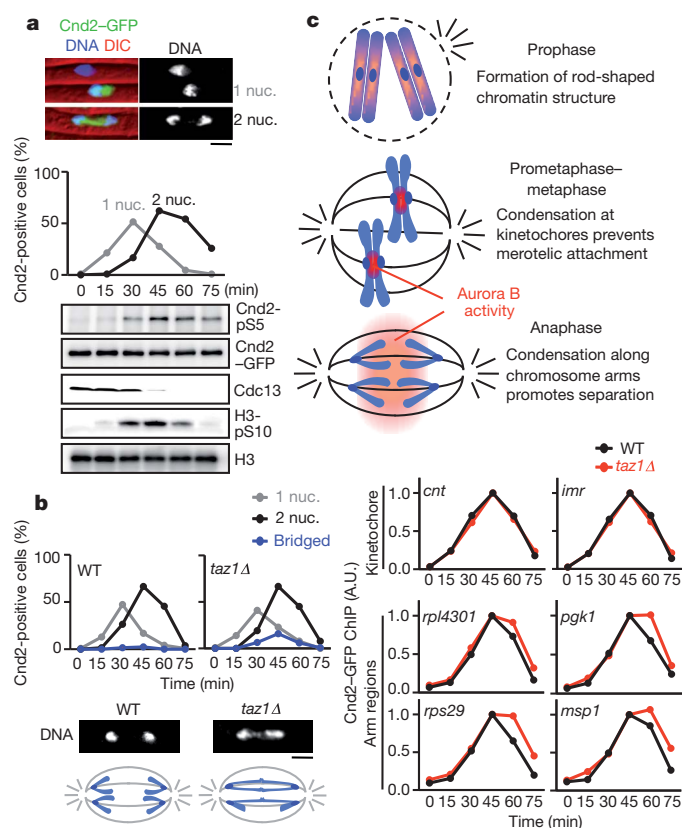


Figure 5 | Spatiotemporal regulation of condensin localization by Aurora B. **a**, Cells expressing Cnd2-GFP were arrested at G2 phase by *cdc25-22* inactivation and released into mitosis. Cell extracts prepared from cells collected at each time point were analysed by immunoblot using the indicated antibodies. Mitotic nuclear division was monitored by nuclear accumulation of Cnd2-GFP and DAPI staining ($n > 100$ cells) (top panels). DIC, differential interference contrast. **b**, Wild-type and *taz1Δ* cells expressing Cnd2-GFP were arrested at G2 phase by *cdc25-22* inactivation, released into mitosis and examined by ChIP assay. A representative anaphase bridge in a *taz1Δ* cell is shown compared with normal anaphase in a wild-type cell. **c**, Schematic depiction of the spatiotemporal regulation of condensin localization by Aurora B. Scale bars, 2 μ m (**a**, **b**).

Discussion

Our results, together with previous findings¹², indicate that fission yeast condensin is enriched at kinetochores and along chromosome arms depending on region-specific recruiters: monopolin homologue Pcs1–Mde4 and Pol III transcription factor TFIIC, respectively. Although recent studies have suggested that the Pcs1–Mde4/Csm1–Lrs1 complex itself functions in crosslinking kinetochore components^{30,31}, our results demonstrate that condensin can replace the complex in this role, at least in fission yeast, indicating that condensin itself may clamp kinetochore chromatin. Budding yeast monopolin, which is required for mono-orientation, may recruit casein kinase 1, rather than condensin, to kinetochores at meiosis I (refs 42, 43), indicating some functional divergence of this complex.

Our study further uncovers that condensin association with chromatin depends largely on the interaction with histone H2A and its variant H2A.Z, which we here define as the chromatin receptor of condensin. The interaction between condensin and the receptor is promoted when Aurora B phosphorylates the condensin subunit Cnd2. The Pcs1–Mde4 pathway and Cnd2 phosphorylation act synergistically to localize condensin at the kinetochores (Fig. 2g). Consistently, although the Pcs1–Mde4 function can be substituted by artificial tethering of Cnd2 to kinetochores, this tethering does not work with the Cnd2-3A protein, which lacks the ability to interact with H2A and H2A.Z (Supplementary Fig. 25). These results support the notion that condensin localization on chromosomes is controlled by two distinct layers of regulation: recruiting and receiving (Fig. 4h). H2A.Z is involved in diverse chromosomal processes such as transcriptional regulation and DNA repair⁴⁴. Similarly, H2A R18, a crucial amino acid for binding with condensin, is implicated in transcriptional silencing and DNA repair⁴⁵, indicating that condensin might influence these processes⁴⁶.

Although previous studies have established that centromeric Aurora B destabilizes erroneous kinetochore–microtubule attachment⁴⁷, our current study suggests that centromeric Aurora B may also promote clamping kinetochore chromatin; both actions thus cooperatively prevent merotelic kinetochore orientation. Our results further indicate that the redistribution of Aurora B at anaphase may promote the phosphorylation of chromatin-unbound condensin around the spindle midzone, which will deliver condensin efficiently to chromosome arms, thereby facilitating their separation (Fig. 5c). These results reaffirm the crucial role of condensin in arm separation at anaphase^{33,48} and explain why longer chromosomes condense more than shorter ones during anaphase⁴⁹.

METHODS SUMMARY

All *Schizosaccharomyces pombe* strains used in this study are listed in Supplementary Table 1. Methods to generate the strains are described in Methods. ChIP assays were conducted as described in Methods. For M phase arrest, we used the *nda3-KM311* mutation or *mad2*⁺ overexpression. For G2/M arrest and release, we used the *cdc25-22* mutant cells cultured at 25 °C, shifted to 36 °C for 4 h to arrest them at G2/M, and released them into mitosis by shifting back to 25 °C. For inactivation of *ark1-as2*, cells were cultured in the presence of 5 μM 1NM-PP1 (1-naphthylmethyl-4-amino-1-tert-butyl-3-(p-methylphenyl)pyrazolo[3,4-d]pyrimidine) at 30 °C. HeLa cells were synchronized by double thymidine block and release, and incubated in the medium containing 330 nM nocodazole. Immunoprecipitation, immunofluorescence and immunostaining were performed as described in Methods. Statistical analysis was done with a two-tailed unpaired *t*-test.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 December 2010; accepted 11 May 2011.

Published online 1 June 2011.

- Hirano, T. Condensins: organizing and segregating the genome. *Curr. Biol.* **15**, R265–R275 (2005).
- Hudson, D. F., Marshall, K. M. & Earnshaw, W. C. Condensin: Architect of mitotic chromosomes. *Chromosome Res.* **17**, 131–144 (2009).
- Koshland, D. & Strunnikov, A. Mitotic chromosome condensation. *Annu. Rev. Cell Dev. Biol.* **12**, 305–333 (1996).

- Ono, T., Fang, Y., Spector, D. L. & Hirano, T. Spatial and temporal regulation of condensins I and II in mitotic chromosome assembly in human cells. *Mol. Biol. Cell* **15**, 3296–3308 (2004).
- Oliveira, R. A., Coelho, P. A. & Sunkel, C. E. The condensin I subunit Barren/CAP-H is essential for the structural integrity of centromeric heterochromatin during mitosis. *Mol. Cell. Biol.* **25**, 8971–8984 (2005).
- Samoshkin, A. *et al.* Human condensin function is essential for centromeric chromatin assembly and proper sister kinetochore orientation. *PLoS ONE* **4**, e6831 (2009).
- Ribeiro, S. A. *et al.* Condensin regulates the stiffness of vertebrate centromeres. *Mol. Biol. Cell* **20**, 2371–2380 (2009).
- Nakazawa, N. *et al.* Dissection of the essential steps for condensin accumulation at kinetochores and rDNAs during fission yeast mitosis. *J. Cell Biol.* **180**, 1115–1131 (2008).
- Nasmyth, K. & Haering, C. H. Cohesin: its roles and mechanisms. *Annu. Rev. Genet.* **43**, 525–558 (2009).
- Gerlich, D., Hirota, T., Koch, B., Peters, J. M. & Ellenberg, J. Condensin I stabilizes chromosomes mechanically through a dynamic interaction in live cells. *Curr. Biol.* **16**, 333–344 (2006).
- Hudson, D. F. *et al.* Molecular and genetic analysis of condensin function in vertebrate cells. *Mol. Biol. Cell* **19**, 3070–3079 (2008).
- Iwasaki, O., Tanaka, A., Tanizawa, H., Grewal, S. I. & Noma, K. Centromeric localization of dispersed Pol III genes in fission yeast. *Mol. Biol. Cell* **21**, 254–265 (2010).
- D'Ambrosio, C. *et al.* Identification of cis-acting sites for condensin loading onto budding yeast chromosomes. *Genes Dev.* **22**, 2215–2227 (2008).
- Johzuka, K. & Horiuchi, T. The cis element and factors required for condensin recruitment to chromosomes. *Mol. Cell* **34**, 26–35 (2009).
- Rabitsch, K. P. *et al.* Kinetochore recruitment of two nucleolar proteins is required for homolog segregation in meiosis I. *Dev. Cell* **4**, 535–548 (2003).
- Takemoto, A. *et al.* The chromosomal association of condensin II is regulated by a noncatalytic function of PP2A. *Nature Struct. Mol. Biol.* **16**, 1302–1308 (2009).
- Liu, W. *et al.* PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. *Nature* **466**, 508–512 (2010).
- Bazile, F., St-Pierre, J. & D'Amours, D. Three-step model for condensin activation during mitotic chromosome condensation. *Cell Cycle* **9**, 3243–3255 (2010).
- Ruchaud, S., Carmena, M. & Earnshaw, W. C. Chromosomal passengers: conducting cell division. *Nature Rev. Mol. Cell Biol.* **8**, 798–812 (2007).
- Giet, R. & Glover, D. M. *Drosophila* aurora B kinase is required for histone H3 phosphorylation and condensin recruitment during chromosome condensation and to organize the central spindle during cytokinesis. *J. Cell Biol.* **152**, 669–682 (2001).
- Hagstrom, K. A., Holmes, V. F., Cozzarelli, N. R. & Meyer, B. J. C. *elegans* condensin promotes mitotic chromosome architecture, centromere organization, and sister chromatid segregation during mitosis and meiosis. *Genes Dev.* **16**, 729–742 (2002).
- Kaitna, S., Pasierbek, P., Jantsch, M., Loidl, J. & Glotzer, M. The Aurora B kinase AIR-2 regulates kinetochores during mitosis and is required for separation of homologous chromosomes during meiosis. *Curr. Biol.* **12**, 798–812 (2002).
- Petersen, J. & Hagan, I. M. S. *pombe* Aurora kinase/survivin is required for chromosome condensation and the spindle checkpoint attachment response. *Curr. Biol.* **13**, 590–597 (2003).
- Lavoie, B. D., Hogan, E. & Koshland, D. *In vivo* requirements for rDNA chromosome condensation reveal two cell-cycle-regulated pathways for mitotic chromosome folding. *Genes Dev.* **18**, 76–87 (2004).
- Takemoto, A. *et al.* Analysis of the role of Aurora B on the chromosomal targeting of condensin I. *Nucleic Acids Res.* **35**, 2403–2412 (2007).
- Lipp, J. J., Hirota, T., Poser, I. & Peters, J. M. Aurora B controls the association of condensin I but not condensin II with mitotic chromosomes. *J. Cell Sci.* **120**, 1245–1255 (2007).
- Mora-Bermúdez, F., Gerlich, D. & Ellenberg, J. Maximal chromosome compaction occurs by axial shortening in anaphase and depends on Aurora kinase. *Nature Cell Biol.* **9**, 822–831 (2007).
- Hsu, J. Y. *et al.* Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes. *Cell* **102**, 279–291 (2000).
- Murnion, M. E. *et al.* Chromatin-associated protein phosphatase 1 regulates aurora-B and histone H3 phosphorylation. *J. Biol. Chem.* **276**, 26656–26665 (2001).
- Corbett, K. D. *et al.* The monopolin complex crosslinks kinetochore components to regulate chromosome-microtubule attachments. *Cell* **142**, 556–567 (2010).
- Gregan, J. *et al.* The kinetochore proteins Pcs1 and Mde4 and heterochromatin are required to prevent merotelic orientation. *Curr. Biol.* **17**, 1190–1200 (2007).
- Yokobayashi, S. & Watanabe, Y. The kinetochore protein Moa1 enables cohesion-mediated monopolar attachment at meiosis I. *Cell* **123**, 803–817 (2005).
- Saka, Y. *et al.* Fission yeast cut3 and cut14, members of a ubiquitous protein family, are required for chromosome condensation and segregation in mitosis. *EMBO J.* **13**, 4938–4952 (1994).
- Haeusler, R. A., Pratt-Hyatt, M., Good, P. D., Gipson, T. A. & Engelke, D. R. Clustering of yeast tRNA genes is mediated by specific association of condensin with tRNA gene transcription complexes. *Genes Dev.* **22**, 2204–2214 (2008).
- Hauf, S. *et al.* Aurora controls sister kinetochore mono-orientation and homolog bi-orientation in meiosis-I. *EMBO J.* **26**, 4475–4486 (2007).
- Koch, A., Krug, K., Pangelley, S., Macek, B. & Hauf, S. Mitotic substrates of the kinase Aurora with roles in chromatin regulation identified through quantitative phosphoproteomics of fission yeast. *Sci. Signal.* doi:10.1126/scisignal.2001588 (in the press).

37. Rangasamy, D., Greaves, I. & Tremethick, D. J. RNA interference demonstrates a novel role for H2A.Z in chromosome segregation. *Nature Struct. Mol. Biol.* **11**, 650–655 (2004).
38. Kim, H. S. *et al.* An acetylated form of histone H2A.Z regulates chromosome architecture in *Schizosaccharomyces pombe*. *Nature Struct. Mol. Biol.* **16**, 1286–1293 (2009).
39. Buchanan, L. *et al.* The *Schizosaccharomyces pombe* JmjC-protein, Msc1, prevents H2A.Z localization in centromeric and subtelomeric chromatin domains. *PLoS Genet.* **5**, e1000726 (2009).
40. Fuller, B. G. *et al.* Midzone activation of aurora B in anaphase produces an intracellular phosphorylation gradient. *Nature* **453**, 1132–1136 (2008).
41. Miller, K. M. & Cooper, J. P. The telomere protein Taz1 is required to prevent and repair genomic DNA breaks. *Mol. Cell* **11**, 303–313 (2003).
42. Brito, I. L., Yu, H. G. & Amon, A. Condensins promote coorientation of sister chromatids during meiosis I in budding yeast. *Genetics* **185**, 55–64 (2010).
43. Petronczki, M. *et al.* Monopolar attachment of sister kinetochores at meiosis I requires casein kinase 1. *Cell* **126**, 1049–1064 (2006).
44. Talbert, P. B. & Henikoff, S. Histone variants—ancient wrap artists of the epigenome. *Nature Rev. Mol. Cell Biol.* **11**, 264–275 (2010).
45. Parra, M. A. & Wyrick, J. J. Regulation of gene transcription by the histone H2A N-terminal domain. *Mol. Cell. Biol.* **27**, 7641–7648 (2007).
46. Aono, N., Sutani, T., Tomonaga, T., Mochida, S. & Yanagida, M. Cnd2 has dual roles in mitotic condensation and interphase. *Nature* **417**, 197–202 (2002).
47. Tanaka, T. U. Kinetochore-microtubule interactions: steps towards bi-orientation. *EMBO J.* **29**, 4070–4082 (2010).
48. Renshaw, M. J. *et al.* Condensins promote chromosome recoiling during early anaphase to complete sister chromatid separation. *Dev. Cell* **19**, 232–244 (2010).
49. Neurohr, G. *et al.* A midzone-based ruler adjusts chromosome compaction to anaphase spindle length. *Science* **332**, 465–468 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Hauf for critically reading the manuscript and K.-i. Noma, T. Hirano and the Yeast Genetic Resource Center (YGRC) for yeast strains or reagents. We also thank S. Hauf and M. Yanagida for communicating unpublished results, and all the members of our laboratory for their support and discussion. This work was supported in part by Global COE Program (Integrative Life Science Based on the Study of Biosignaling Mechanisms) (to K.T. and H.S.), Special Coordination Funds for Promoting Science and Technology (to T.S.) and a Grant-in-Aid for Specially Promoted Research, MEXT, Japan (to Y.W.).

Author Contributions K.T., supported by T.S., performed most of the experiments using fission yeast cells and proteins. H.S. performed all experiments using human cells or proteins. Experimental design and interpretation of data were conducted by all authors. Y.W. supervised the project and K.T. and Y.W. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.W. ([ywatana@iam.u-tokyo.ac.jp](mailto:ywatanab@iam.u-tokyo.ac.jp)).

METHODS

Schizosaccharomyces pombe strain. For the construction of TEV protease-cleavable Cnd2, the target site for TEV protease (Glu-Asn-Leu-Tyr-Phe-Gln-Gly(-Ala-Ser)⁵⁰) was inserted between Val (amino acid 376) and Val (amino acid 377) in Cnd2-6HA (six copies of HA, abbreviated as HA hereafter) by site-directed mutagenesis. The resulting *cnd2(TEV)*-HA or control *cnd2*-HA construct carrying the 3'-untranslated region and the *kan^r* marker within it was integrated into the chromosome, replacing the endogenous *cnd2⁺* gene. To express Cnp3C-CFP-TEV, CFP-TEV, or Cnp3C-CFP protein³², the sequence encoding them was cloned under the promoter *Padh21* (a weak version of the *adh1⁺* promoter) and integrated into the locus adjacent to the *zfs1⁺* gene of chromosome 2 (denoted by the Z locus) using the *nat^r* marker. Deletion of *pcs1⁺*, *cnd2⁺*, *hta1⁺*, *hta2⁺*, *taz1⁺* and *sgo2⁺*, tagging of endogenous *mde4⁺*, *cnd2⁺*, *ark1⁺* and *cut3⁺* by Myc, GFP, or HA, and promoter exchange of endogenous *phl1⁺* by *Pnmt81* (a weak version of the *nmt1⁺* promoter) were performed using the PCR-based gene targeting method for *S. pombe*⁵¹. To express Cnd2-tdTomato-Cnp3C, a sequence encoding tdTomato and Cnp3C (amino acids 384–642) was fused to the carboxy terminus of Cnd2 and cloned under the promoter *Padh31* (a weak version of the *adh1⁺* promoter). The resulting plasmid was linearized and integrated into the locus adjacent to the SPAC26F1.12c gene of chromosome 1 (denoted by the C locus) using the *hyg^r* marker. To generate the *cnd2*-3A strain, S5, S41 and S52 were changed to alanines using Kunkel methods. The genomic *cnd2* fragments carrying the mutations were then transformed into *cnd2::ura4⁺* cells (additional copy of *cnd2⁺* under the promoter *Padh15* was integrated at the C locus with *hyg^r* marker) and the integration at *cnd2::ura4⁺* was selected by 5-fluoroorotic acid (5-FOA) resistance and confirmed by PCR. The *cnd2*-3E strain was similarly constructed. For the analysis of Cnd2-3A localization, wild-type *cnd2* or *cnd2*-3A, fused with GFP, was cloned under the promoter *Padh21* and integrated into the Z locus using the *nat^r* marker. For generating the *h2a-R18A* strain, R18 of *hta1⁺* and *hta2⁺* were changed to alanines, using the PrimeSTAR Mutagenesis Basal Kit (Takara). The genomic *hta1* and *hta2* fragments carrying the mutations were then transformed into *hta1::ura4⁺* and *hta2::ura4⁺* cells, respectively, and integrants were selected for 5-FOA resistance and confirmed by PCR. After crossing *hta1-R18A* and *hta2-R18A* cells, double mutants (*h2a-R18A*) were selected. The *h2a-R18Q* or *h2a-R18K* strain was similarly constructed. To visualize tubulin, a sequence encoding mCherry or CFP was fused to the N terminus of *atb2⁺*, cloned under the promoter *Padh13* or *Padh15* (a weak version of the *adh1⁺* promoter), and integrated into the Z or C locus using *nat^r* or *hyg^r* marker. To visualize nucleolus, Gar1-CFP (ref. 52) was expressed from the *Padh21* promoter (integrated at C locus using *hyg^r* marker). All strains used are listed in Supplementary Table 1.

Introduction of a *lacO* or *tetO* array into various chromosomal regions. A *lacO* array engineered from pCT31 (a gift from A. Yamamoto) or a *tetO* array engineered from p306*tetO*_{x112} (a gift from K. Nasmyth) was used in this study. For the visualization of *cnt1* and *dh1* (ref. 53), *lacO* arrays were introduced into these two regions simultaneously. To obtain *tel-lacO* and *arm-lacO* (centromeric side of the rDNA locus), a fragment containing a *lacO* array with the *kan^r* cassette, flanked by the regions adjacent to the telomeric region at the left arm of chromosome 1 (9282–9805 and 9806–10254 of chromosome 1) or to the centromeric side of the rDNA locus at right arm of chromosome 3 (2428525–2429283 and 2429284–2430083 of chromosome 3), was used for transformation, respectively. To obtain *imr1-tetO* and *imr3-tetO*, fragments containing *tetO* with the *ura4⁺* cassette, flanked by the *imr1* regions (3764007–3764631 and 3764632–3765171 of chromosome 1) or the *imr3* regions (1093757–1094520 and 1094521–1095451 of chromosome 3), were used for transformation, respectively. Correct integration was confirmed by PCR. Among symmetric sequences of *dh1*, *imr1*, or *imr3*, the integrated locus (left or right) was also determined by PCR.

Two-hybrid assay. The constructs of *Pcs1* (amino acids 1–222) and *Mde4* (amino acids 1–421) were amplified by PCR and cloned into pGBKT7 vectors and used as bait. The constructs of *Cut14* (1–1172), *Cut3(N)* (1–477), *Cut3(M)* (468–882), *Cut3(C)* (873–1324), *Cnd2(N)* (1–250), *Cnd2(M)* (241–500), *Cnd2(C)* (491–742), *Cnd1(N)* (1–600), *Cnd1(C)* (591–1158) and *Cnd3* (1–875) were amplified by PCR and cloned into pGADT7 vectors and used as prey. These plasmids were transformed into *Saccharomyces cerevisiae* AH109 strain. Plates lacking histidine and adenine and containing 2 mM 3AT (3-aminotriazole) were used as selective media.

Culture of fission yeast. For prometaphase arrest, we used the *nda3-KM311* mutation⁵⁴ and cultured cells at 17 °C for 13 h. For metaphase arrest, we used the pREP1(*mad2⁺*) plasmids expressed by the *nmt1⁺* promoter (gift from T. Matsumoto) and cultured cells in the minimal medium at 30 °C for 14–16 h. For inactivation of *ark1-as2*, cells were cultured in the presence of 5 μM 1NM-PP1 (1-naphthylmethyl-4-amino-1-tert-butyl-3-(*p*-methylphenyl)pyrazolo[3,4-*d*]pyrimidine) at 30 °C. For the inactivation of *sfc3-1*, cells were cultured at

34 °C for the last 2 h of arrest. For the depletion of H2A.Z by the *Pnmt81* promoter shut off, we cultured cells in YE medium (+ thiamine) for 24 h at 25 °C. For G2/M arrest and release, we used the *cdc25-22* mutant cells cultured at 25 °C, shifted to 36 °C for 4 h to arrest them at G2/M, and released them into mitosis by shifting back to 25 °C. For the induction of telomere entanglement in *taz1Δ* cells, cells were cultured at 20 °C for 5 h just before shifting to 36 °C.

Preparation of phospho-specific antibodies. To generate phospho-specific antibodies against Cnd2-pS5, phosphorylated peptides (MKRA(pS)LGGH) were used to immunize rabbits (Operon). Antibodies were purified from anti-serum with phosphorylated peptide-conjugated CNBr-activated sepharose and dialysed against PBS. Similarly, for the production of phospho-specific antibodies against hCAP-H-pS70, phosphorylated peptides (LQRRR(pS)RVFDL) were used to immunize rabbits (Operon).

Chromatin immunoprecipitation assay. The procedure was carried out essentially as described previously⁵². Anti-HA antibodies (12CA5, Roche), anti-GFP antibodies (Living Colours Full-length A.v. Polyclonal Antibody, Clontech), anti-Myc antibodies (9E10, Santa Cruz), anti-H3-pS10 antibodies (Upstate), anti-Cnp1 antibodies⁵⁵ and anti-Psc3 antibodies⁵⁶ were used for immunoprecipitation. DNA prepared from whole-cell extracts or immunoprecipitated fractions was analysed by quantitative PCR with the ABI PRISM7000 system (Applied Biosystems) using SYBR Premix Ex Taq (Perfect Real Time) (Takara). The primers used for PCR were described previously^{32,52,57} except those for *dg* (5'-CACATCATCGTCGTAACAT-3' and 5'-GATATCATCTATATTTAATGACTACT-3'), *dh* (5'-GAAACACATCGTTGTCTTCAGAG-3' and 5'-CGTCTGTAGCTGCATGTGAA-3'), *rpl4301* (5'-GTATGGAAGCAAGGTA-3' and 5'-GATGGTGTATAAACTTTAGT-3'), *rps29* (5'-CGTAACTTGGTATGAAAGCA-3' and 5'-TTAATTACAGTAATCACAC-3'), *pgk1* (5'-TCAATAAGTTGGATCTTGTC-3' and 5'-CTTATGAAGAAGGAGCTCGA-3') and *rDNA-N1* (5'-GAAAAAGTCGAGCGAGTCG-3' and 5'-TCAAAAGTGCA TTACCCTTAC-3'). We included control IgG or untagged immunoprecipitation in each experiment to account for non-specific binding in the ChIP fractions. We normalized the immunoprecipitation efficiency among samples by performing Cnp1 immunoprecipitation. Error bars indicate s.d. of three PCR amplifications. We confirmed that all ChIP data are reproducible in separate experiments.

Co-immunoprecipitation from fission yeast extracts. Cultured cells were harvested, suspended in HB buffer (25 mM MOPS (pH 7.2), 15 mM MgCl₂, 15 mM EGTA, 60 mM β-glycerophosphate, 0.1 mM Na-orthovanadate, 0.1 mM NaF, 15 mM *p*-nitrophenylphosphate, 1% Triton X-100, 1 mM dithiothreitol, 1 mM PMSF, complete protease inhibitor (Roche)), and disrupted using the Multi-bead shaker (Y. Kikai). Then the supernatants, collected after centrifugation, were incubated with anti-Myc antibody (9E10, Santa Cruz) or anti-GFP polyclonal antibodies (Living Colours Full-length A.v. Polyclonal Antibody, Clontech) for 1.5 h at 4 °C. Protein A beads (Amersham) were added and incubation was continued for 1.5 h at 4 °C. After washing with HB buffer, we analysed the immunoprecipitates by SDS-PAGE and western blotting with anti-HA (1:1,000, 12CA5, Roche), anti-GFP (1:1,000, Roche), anti-Myc (1:1,000, 9E10, Santa Cruz), anti-Cnd2-pS5 (1:1,000), anti-Cdc13 (1:1,000), anti-H3 (1:1,000, Abcam), anti-H3-pS10 (1:1,000, Upstate), anti-Psc3 (ref. 56) (1:1,000) and TAT1 (1:5000) antibodies.

In vitro kinase and pull-down assays of fission yeast proteins. The constructs of Ark1 (amino acids 1–355), Pic1IN-box (914–1018), Cut14(N) (1–620), Cut14(C) (611–1172), Cut3(N) (1–477), Cut3(M) (468–882), Cut3(C) (873–1324), Cnd2(N) (1–250), Cnd2(M) (241–500), Cnd2(C) (491–742), Cnd1(N) (1–600), Cnd1(C) (591–1158) and Cnd3 (1–875) were amplified by PCR and cloned into pGEX4T-1, pGEX4T-2, or pGEX4T-3 vectors. For kinase assay, kinases and substrates were incubated with kinase buffer (50 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 1% Triton X-100) at 30 °C for 30 min in the presence of [γ -³²P]ATP. Incorporation of the radioactive phosphate group was visualized by autoradiography and protein loading was analysed by staining with Coomassie blue (CBB). For pull-down assay, MBP and MBP-fused proteins were purified from *E. coli* and bound to amylose beads (NEB). GST-fused proteins, pre-incubated with GST-Ark1 and GST-Pic1IN in kinase buffer (50 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 0.1% Triton X-100) at 30 °C for 30 min with or without ATP, were mixed with beads and incubated in binding buffer (20 mM Tris-HCl (pH 7.5), 1 mM EDTA, 200 mM NaCl, 10% glycerol, 0.8% NP-40, 1 mg ml⁻¹ BSA, Complete protease inhibitor (Roche)) at 4 °C for 1.5 h. After washing with binding buffer, the bound proteins were analysed by western blotting with anti-GST antibodies (1:500, Santa Cruz). Protein loading was analysed by staining with CBB.

Time-lapse imaging. For the measurement of the distance between *imr3*-tdTomato signals during anaphase (Supplementary Fig. 5), wild-type, *cut3-477* or *pcs1Δ* cells were cultured in minimal medium at 30 °C for 4 h, then live cell recordings were performed on a DeltaVision P system (Applied Precision) in an

air-conditioned chamber maintained at 30 °C. A glass-bottomed dish (Matsunami) coated with 0.2% concanavalin A (Sigma) was used to mount cells. During observation, the cells transferred to the glass-bottomed dish were supplied with minimal medium. Images were acquired by Z-sectioning with 45-s intervals, and deconvoluted in the softWoRx software (Applied Precision). To measure the distance of *imr3*-tdTomato signals, softWoRx software (Applied Precision) was used.

Quantification of fluorescent signals in fission yeast. To quantify the fluorescent signals at centromeres, in-focus images of Ark1-GFP cells were taken with Axio Vision imaging software (Carl Zeiss). We measured the average intensity of the centromeric dots and subtracted the average background intensity.

Culture, synchronization and RNAi of HeLa cells. HeLa cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS, 0.03% L-glutamine, 100 units ml⁻¹ penicillin, and 100 mg ml⁻¹ streptomycin in a moist atmosphere. Nocodazole (Sigma) and thymidine (Sigma) were added at 330 nM and 2 mM, respectively, for synchronization. RNAi was performed as described⁵⁸. Sequences of siRNA against *hH2A.Z*, 5'-GAAGAAAGGACAACA GAAGdTdT-3' and siRNA against *BubR1*, 5'-CGGGCAUUGAAUAUGAA AdTdT-3'.

Expression of GFP-hCAP-H. Full-length hCAP-H cDNA was cloned from HeLa RNA by PCR. Full-length wild type and hCAP-H-S70A tagged with GFP at their N termini were inserted into pcPURO vectors derived from Y. Goto's laboratory for transient expression. Vectors were transfected into HeLa cells using Lipofectamine LTX (Invitrogen).

In vitro kinase and pull-down assays of human proteins. GST-fused and MBP-fused proteins were bacterially expressed and purified with glutathione-sepharose beads (GE) and amylose resin (NEB), respectively. For kinase assays, reactions were carried out in 30 µl of kinase buffer (50 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 0.5% Triton X-100 supplemented with Complete protease inhibitor (Roche), 1 mM dithiothreitol, 1 mM PMSF) in the presence of [γ -³²P]ATP. Incorporation of radioactive phosphate group was visualized by autoradiography and protein level was analysed by staining with CBB. For pull-down assay, purified proteins were phosphorylated with GST-Aurora B in kinase buffer in the presence of 0.25 mM ATP at 30 °C for 45 min. Phosphorylated proteins were mixed with MBP-tagged proteins beads and incubated in binding buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% glycerol, 0.8% NP40 supplemented with Complete protease inhibitor, 1 mM dithiothreitol, 1 mg ml⁻¹ BSA) at 4 °C for 3 h. The beads were then washed five times with binding buffer. The bead-bound proteins were subjected to SDS-PAGE and transferred onto the Immunobilon PVDF membrane (Millipore) and detected by immunoblot using anti-GST (1:500, Santa Cruz). Protein loading was analysed by staining with CBB.

Preparation of HeLa cell extracts and immunoprecipitation. HeLa cells were synchronized by single thymidine block and release, and incubated in the medium containing nocodazole for 16 h. Mitotic cells were collected by mitotic shake-off and washed with PBS once, then re-suspended in CSK buffer (10 mM PIPES pH 6.8, 10% glycerol, 3 mM MgCl₂, supplemented with Complete protease inhibitor, PhosSTOP phosphatase inhibitor mixture (Roche), 1 mM dithiothreitol, 0.25 mM PMSF) containing 100 mM NaCl. After incubation at 4 °C for 30 min, supernatants (chromatin-unbound) were extracted and the insoluble chromatin pellet was collected by centrifugation at 16,100g for 10 min, washed once with PBS and re-suspended in CSK buffer containing 250 mM NaCl, 1 mM CaCl₂ and

digested with micrococcal nuclease (Sigma) at 30 °C for 15 min. After centrifugation at 16,100g for 10 min, supernatants (chromatin-bound) were prepared. Whole-cell extracts (WCE)—the mixture of chromatin-bound and chromatin-unbound fractions—were immunoblotted with the following antibodies: anti-hCAP-H-pS70 (1:3000), anti-H3-pS10 (1:1,000, Cell Signaling), anti-cyclin B (1:1,000, Santa Cruz), anti-tubulin (1:5,000, DM1A; Sigma), anti-Aurora B (1:500, BD Bioscience), anti-GFP (1:1,000, Roche). The chromatin-bound fraction was rotated with anti-SMC2 (Abcam) or rabbit control IgG for 1 h at 4 °C, followed by 1.5 h rotation with protein A sepharose beads (GE) at 4 °C. The beads were washed five times with CSK buffer containing 250 mM NaCl. We analysed the immunoprecipitates by silver staining using the SilverQuest Staining kit (Invitrogen) or by immunoblot with anti-SMC2 (1:5,000), anti-CAP-H (1:2,000, Bethyl), anti-H2A (1:1,000, Abcam), anti-H2B (1:1,000, Millipore), anti-H3 (1:1,000, Abcam), anti-H2A.Z (1:1,000, Cell Signaling) and anti-actin (1:1,000, Santa Cruz) antibodies.

Preparation of ZM447439-treated cell extracts. HeLa cells were synchronized by double thymidine block and release, and incubated in medium containing 1,000 nM nocodazole for 14 h in the presence of 4 µM ZM447439 during the last 2 h. Mitotic cells were collected by mitotic shake-off.

Immunostaining of HeLa cells. Immunofluorescence staining was performed as described⁵⁸ with anti-GFP (1:1,000, Invitrogen), anti-H3-pS10 (1:1,000), anti-CAP-H (1:500), biotinylated anti-CAP-G (1:500, a gift from T. Hirano), anti-H2A.Z (1:200) antibodies. DNA was stained with 3 mg ml⁻¹ Hoechst 33342 (Wako). For pre-extraction, the cells were permeabilized in 0.1% Triton X-100 in XBE2 (10 mM HEPES pH 7.7, 2 mM MgCl₂, 100 mM KCl and 5 mM EGTA) at room temperature for 1 min and then fixed with 2% paraformaldehyde in XBE2 for 15 min. Immunostaining images were obtained with the use of DeltaVision softWoRx software (Applied Precision). Image segmentation into chromatin and cytoplasm was done using the Hoechst signal. Average cytoplasm signals were subtracted from chromatin signal and normalized to the Hoechst signal as reported²⁶. Signal intensities of Aurora B, CAP-H and DNA were measured using the ImageJ program.

50. Dougherty, W. G., Cary, S. M. & Parks, T. D. Molecular genetic analysis of a plant virus polypeptide cleavage site: a model. *Virology* **171**, 356–364 (1989).
51. Bähler, J. *et al.* Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* **14**, 943–951 (1998).
52. Yokobayashi, S., Yamamoto, M. & Watanabe, Y. Cohesins determine the attachment manner of kinetochores to spindle microtubules at meiosis I in fission yeast. *Mol. Cell. Biol.* **23**, 3965–3973 (2003).
53. Sakuno, T., Tada, K. & Watanabe, Y. Kinetochore geometry defined by cohesion within the centromere. *Nature* **458**, 852–858 (2009).
54. Hiraoka, Y., Toda, T. & Yanagida, M. The *NDA3* gene of fission yeast encodes β -tubulin: a cold-sensitive *nda3* mutation reversibly blocks spindle formation and chromosome movement in mitosis. *Cell* **39**, 349–358 (1984).
55. Takayama, Y. *et al.* Biphasic incorporation of centromeric histone CENP-A in fission yeast. *Mol. Biol. Cell* **19**, 682–690 (2008).
56. Nonaka, N. *et al.* Recruitment of cohesin to heterochromatic regions by Swi6/HP1 in fission yeast. *Nature Cell Biol.* **4**, 89–93 (2002).
57. Kawashima, S. A. *et al.* Shugoshin enables tension-generating attachment of kinetochores by loading Aurora to centromeres. *Genes Dev.* **21**, 420–435 (2007).
58. Kitajima, T. S. *et al.* Shugoshin collaborates with protein phosphatase 2A to protect cohesin. *Nature* **441**, 46–52 (2006).

Induction of human neuronal cells by defined transcription factors

Zhiping P. Pang^{1*}, Nan Yang^{2*}, Thomas Vierbuchen^{2,3*}, Austin Ostermeier^{2,3}, Daniel R. Fuentes², Troy Q. Yang², Ami Citri⁴, Vittorio Sebastiano², Samuele Marro², Thomas C. Südhof^{1,5} & Marius Wernig^{2,3}

Somatic cell nuclear transfer, cell fusion, or expression of lineage-specific factors have been shown to induce cell-fate changes in diverse somatic cell types^{1–12}. We recently observed that forced expression of a combination of three transcription factors, *Brn2* (also known as *Pou3f2*), *Ascl1* and *Myt1l*, can efficiently convert mouse fibroblasts into functional induced neuronal (iN) cells¹³. Here we show that the same three factors can generate functional neurons from human pluripotent stem cells as early as 6 days after transgene activation. When combined with the basic helix–loop–helix transcription factor *NeuroD1*, these factors could also convert fetal and postnatal human fibroblasts into iN cells showing typical neuronal morphologies and expressing multiple neuronal markers, even after downregulation of the exogenous transcription factors. Importantly, the vast majority of human iN cells were able to generate action potentials and many matured to receive synaptic contacts when co-cultured with primary mouse cortical neurons. Our data demonstrate that non-neural human somatic cells, as well as pluripotent stem cells, can be converted directly into neurons by lineage-determining transcription factors. These methods may facilitate robust generation of patient-specific human neurons for *in vitro* disease modelling or future applications in regenerative medicine.

Encouraged by our recent findings in mouse cells¹³, we explored in this study whether human fibroblasts could also be directly induced to become functional neurons. This was unclear given the differences in the gene regulatory networks governing human and rodent neural development^{14–16}. First, we sought to determine whether forced expression of transcription factors could induce a neuronal fate in human pluripotent cells. To that end, we infected undifferentiated human embryonic stem (ES) cells in chemically defined N3 media¹⁷ with *Brn2*, *Ascl1* and *Myt1l* (BAM) using doxycycline (dox)-inducible lentiviral vectors together with an EGFP (enhanced green fluorescent protein) virus. The majority of ES cells were EGFP-positive 24 h after addition of doxycycline (Supplementary Fig. 1). Strikingly, as early as 3 days after doxycycline treatment, we observed bipolar neuron-like cells surrounding nearly all ES cell colonies (Fig. 1a and Supplementary Fig. 1). By day 8, cells with more mature neuronal morphologies that expressed both β -III-tubulin (Tuj1, also known as TUBB3) and MAP2 had migrated away from ES cell colonies and were present throughout the plate (Fig. 1b, c). In contrast, after infection with EGFP virus alone, no neuronal cells were generated during the same timeframe, and nearly all ES cells had died due to the cytosine β -D-arabinofuranoside (Ara-C) treatment. We then determined the relative contribution of the three factors and found that *Ascl1* alone was sufficient to induce MAP2-positive cells (Supplementary Fig. 2). The addition of *Brn2* or *Myt1l* or both did not increase the efficiency of neuronal differentiation but induced more complex morphologies. Cells infected with all three factors together had the most mature neuronal morphologies

(Supplementary Fig. 2). Electrophysiological analysis surprisingly revealed that as early as 6 days after induction all recorded cells ($n = 16$) generated action potentials (Fig. 1d, e). At day 15 after doxycycline treatment, the average resting membrane potential of neuronal cells was -51 ± 1.8 mV (mean \pm s.e.m., $n = 18$) (Fig. 1f

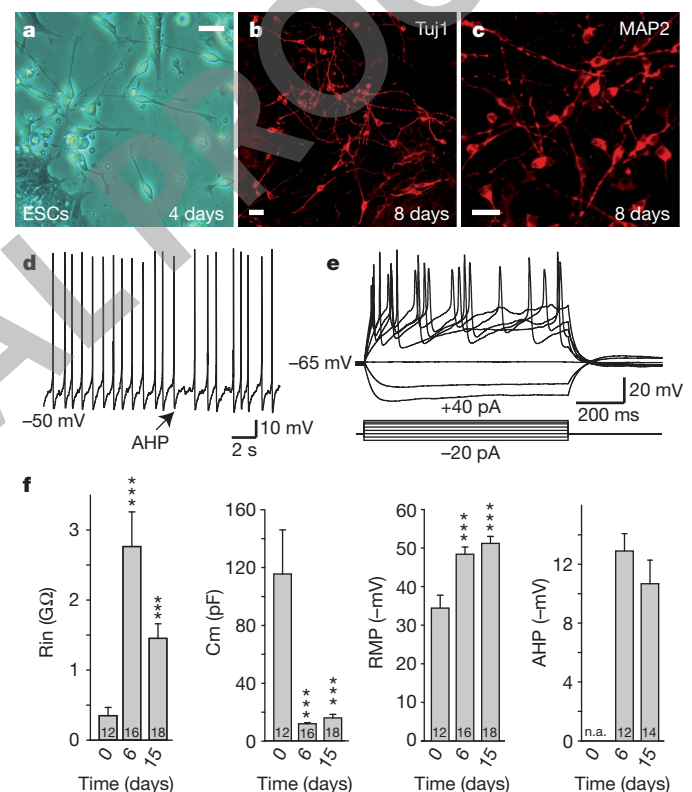


Figure 1 | Rapid generation of functional neurons from human ES cells.

a, Four days after induction, ES-iN cells showed bipolar neuronal morphologies. b, c, Eight days after induction, ES-iN cells expressed Tuj1 (b) and MAP2 (c). d, Spontaneous action potentials presumably caused by membrane potential fluctuations recorded from an ES-iN cell 6 days after induction. Arrow, pronounced afterhyperpolarization potentials (AHP). e, Representative traces of action potentials in response to step current injections 15 days after induction. Membrane potential was maintained at approximately -63 mV. f, Quantification of intrinsic membrane properties in control ES cells (0 day) before and after viral transduction. Rin, membrane input resistance; RMP, resting membrane potential; Cm, capacitance. Scale bars, 10 μ m (a, b, c). Numbers of cells recorded are labelled in the bars. Note the heterogeneity of the parameters (see also Supplementary Fig. 1). Data are presented with mean \pm s.e.m. *** $P < 0.05$.

¹Department of Molecular and Cellular Physiology, Stanford University School of Medicine, 265 Campus Drive, Stanford, California 94305, USA. ²Institute for Stem Cell Biology and Regenerative Medicine, Department of Pathology, Stanford University School of Medicine, 265 Campus Drive, Stanford, California 94305, USA. ³Program in Cancer Biology, Stanford University School of Medicine, 265 Campus Drive, Stanford, California 94305, USA. ⁴Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, 265 Campus Drive, Stanford, California 94305, USA. ⁵Howard Hughes Medical Institute, Stanford University School of Medicine, 265 Campus Drive, Stanford, California 94305, USA.

*These authors contributed equally to this work.

and Supplementary Table 1). These ES-iN cells exhibited prominent after-hyperpolarization potentials (AHPs) following action potentials (Fig. 1d, f). Similar findings could be observed when human induced pluripotent stem (iPS) cells were infected (Supplementary Fig. 3). Thus, the BAM factors rapidly induce neuronal differentiation of human pluripotent stem cells.

Next, we asked whether human fibroblasts could also be directly converted into neurons. To this end, we derived three independent primary human fetal fibroblast lines (HFFs) (see Methods) and performed an extensive characterization of these cultures in various growth conditions to confirm that they lack spontaneous neuronal differentiation potential and do not contain detectable amounts of neural crest stem cells (see Supplementary Fig. 4). Strikingly, 7–10 days after infection with the BAM factors we could detect cells with immature neuronal morphologies. These cells expressed *Tuj1* (Supplementary Fig. 5a), but remained functionally immature as revealed by their inability to generate action potentials 20 days after doxycycline treatment (Supplementary Fig. 5b). Thus, the BAM factors seemed to induce neuronal features but were insufficient to generate functional neurons from human fetal fibroblasts under these conditions.

Therefore, we screened 20 additional factors that could improve the generation of neuronal cells in combination with the BAM pool. We observed that *NeuroD1*, another basic helix–loop–helix transcription factor, improved the efficiency of generating *Tuj1*-positive neuronal cells two to threefold after 3 weeks (Fig. 2a). To determine the relative contribution of *NeuroD1*, we tested various combinations of these four factors. *NeuroD1* alone had no effect, but surprisingly in combination with *Brn2* it was sufficient to generate a similar number of *Tuj1*-positive neuronal cells compared to the BAN (*Brn2*, *Ascl1* and *NeuroD1*), BMN (*Brn2*, *Myt1l* and *NeuroD1*) and BAMN (*Brn2*, *Ascl1*, *Myt1l* and *NeuroD1*) pools (Supplementary Fig. 6a). However, further morphological and functional characterization showed that the BAMN combination generated the most mature neuronal cells (Supplementary Fig. 6b). We therefore decided to focus the further analysis on BAMN-iN cells.

Two weeks after induction, BAMN-iN cells showed neuronal morphologies and were labelled with pan-neuronal antibodies such as anti-*Tuj1*, anti-*NeuN* (also known as *Rbfox3*), anti-PSA-NCAM (polysialylated neural cell adhesion molecule) and anti-MAP2 (Fig. 2b–f). After extended culture periods of 4–5 weeks, we could detect cells expressing neurofilaments (Supplementary Fig. 7a), and rare neuronal processes decorated with punctate staining of synapsin and synaptotagmin, two synaptic vesicle proteins (Fig. 2g, h and Supplementary Fig. 7b). To ensure the co-expression of pan-neuronal and subtype specific markers, we performed single-cell gene-expression profiling of iN cells using Fluidigm dynamic RT-PCR arrays¹⁸. We analysed 54 single HFF-iN cells 34 days after doxycycline treatment from two independently infected cultures (Supplementary Fig. 8). These data revealed robust co-expression of multiple pan-neuronal and synaptic markers in 50/54 HFF-iN cells (β -III-tubulin, *DCX*, *MAP2*, *NCAM*, synapsin). Over half (29 of 54) of the iN cells analysed expressed mRNAs typical for glutamatergic neurons, such as *vGLUT1*, *vGLUT2* (also known as *SLC17A7* and *SLC17A7*, respectively) or both (Fig. 2i). Only two iN cells expressed *GAD67* (also known as *Gad1*) in the absence of *vGLUT1* or 2, and no iN cell expressed the inhibitory marker gene *vGAT* (also known as *SLC32A1*). Five cells expressed the catecholaminergic neuron marker tyrosine hydroxylase. Immunofluorescence analysis revealed that 5 weeks after infection $17 \pm 8\%$ of iN cells expressed the forebrain marker *Tbr1*, $21 \pm 9\%$ expressed the marker of peripheral neurons peripherin, whereas *En1*, a marker of midbrain neurons, serotonin and choline acetyltransferase were not detectable (Supplementary Fig. 9).

To assess whether the iN cell state was stable without continued transgene expression, we monitored the mRNA expression levels of endogenous and exogenous BAMN genes before and after doxycycline induction and after doxycycline withdrawal. Whereas the exogenous

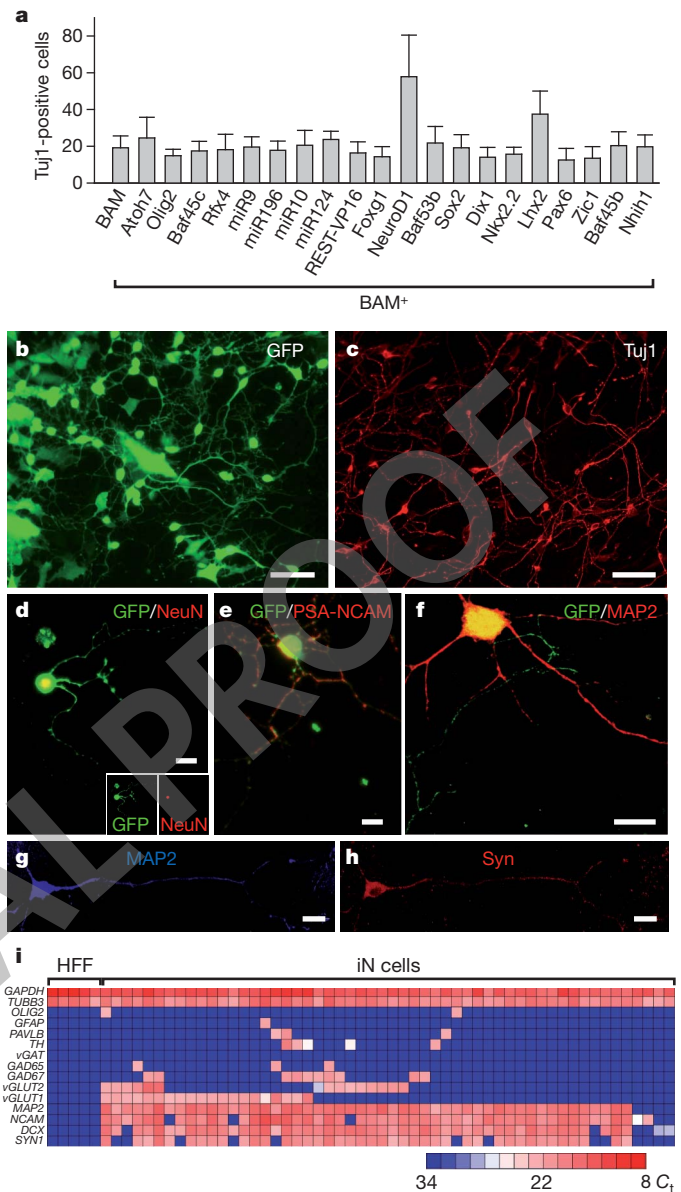


Figure 2 | *NeuroD1* increases reprogramming efficiency in primary human fetal fibroblasts. **a**, Quantification of *Tuj1*-positive BAM-iN cells with indicated factors, 3 weeks after doxycycline treatment. **b**, **c**, Three weeks after doxycycline treatment BAM⁺*NeuroD1* iN cells showed neuronal morphologies (**b**) and expressed *Tuj1* (**c**). **d**–**f**, iN cells expressed NeuN (**d**), PSA-NCAM (**e**), and MAP2 (**f**) 2 weeks after doxycycline treatment. **g**, **h**, An iN cell expressing MAP2 (**g**) and synapsin (**h**) 4 weeks after doxycycline treatment and co-cultured with primary astrocytes. **i**, Single-cell gene expression profiling using Fluidigm dynamic arrays. Rows represent the evaluated genes and columns represent individual cells. Heat map (blue to red) represents the threshold C_t values as indicated. Data in (a) are presented as mean \pm s.d. Scale bars, 100 μ m (**b**, **c**), 10 μ m (**d**–**h**).

transgenes were clearly doxycycline-dependent, the four corresponding endogenous genes were rapidly induced and exhibited increasing expression levels over time even after doxycycline withdrawal (Supplementary Fig. 10). Similarly, HFF-iN cells could be maintained in the absence of doxycycline for 3 weeks (Supplementary Fig. 11).

We next asked whether iN cells generated from HFFs had active membrane properties. iN cells were identified by EGFP fluorescence (Fig. 3b) and whole-cell recordings were performed 14–35 days after doxycycline treatment. Two to three weeks after addition of doxycycline the average resting membrane potential of HFF-iN cells was

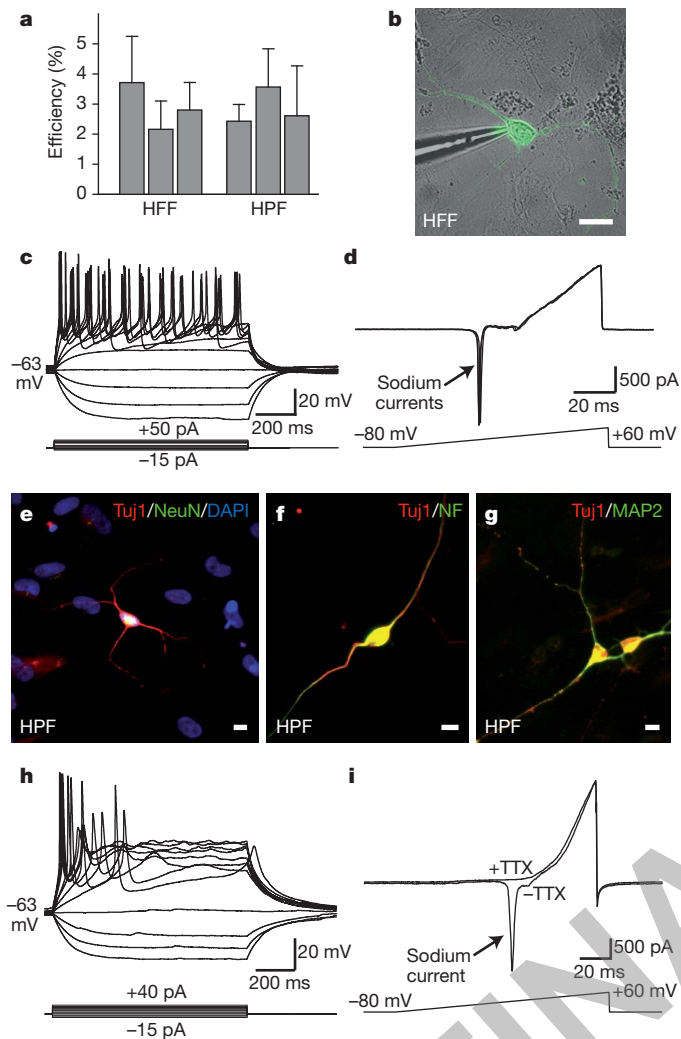


Figure 3 | Membrane properties of fibroblast iN cells. **a**, Quantification of TuJ1-positive neuronal cells from HFFs (line HFF-A) 3 weeks after doxycycline treatment or HPFs (line HPF-B) 4 weeks after doxycycline treatment. $n = 3$ independent experiments. **b**, Patch clamp recording was conducted on HFF-iN cells identified by EGFP fluorescence and differential interference contrast microscopy. **c**, Representative traces of membrane potentials in response to step current injections (lower panel) from an HFF-iN cell 19 days after doxycycline treatment. Membrane potential was maintained at approximately -63 mV. **d**, Representative traces of membrane currents recorded with a ramp protocol (lower panel). Fast activating and inactivating Na^+ currents were prominent. Three traces are shown superimposed. **e–g**, HPF-iN cells express TuJ1 (red) and NeuN (green) (**e**), neurofilament (green, NF) (**f**) and MAP2 (green) (**g**). **h**, Representative traces of membrane potentials in response to step current injections in HPF-iN cells. Action potentials were generated in cultures without glia. **i**, Representative traces of membrane currents recorded following a ramp protocol (lower panel) in HPF-iN cells. The Na^+ currents could be blocked by tetrodotoxin (TTX). Data in (**a**) are presented as mean \pm s.d. Scale bars, 10 μm (**a**, **e–g**).

-52.2 ± 2.2 mV (mean \pm s.e.m., $n = 41$). When HFF-iN cells were step-depolarized, action potentials could be detected in many iN cells at 14–25 days, and in all recorded iN cells at days 34–35 (Fig. 3c, Supplementary Figs 7 and 12 and Supplementary Table 1). Fast-activating and inactivating inward Na^+ currents as well as outward K^+ currents were also observed (Fig. 3d).

To determine whether the BAMN factors were also capable of converting more mature human fibroblasts into iN cells, we derived primary human postnatal fibroblasts (HPFs) from three different perinatal foreskin resections. In all three HPF lines, expression of BAMN factors reproducibly generated neuron-like cells with co-expression of multiple

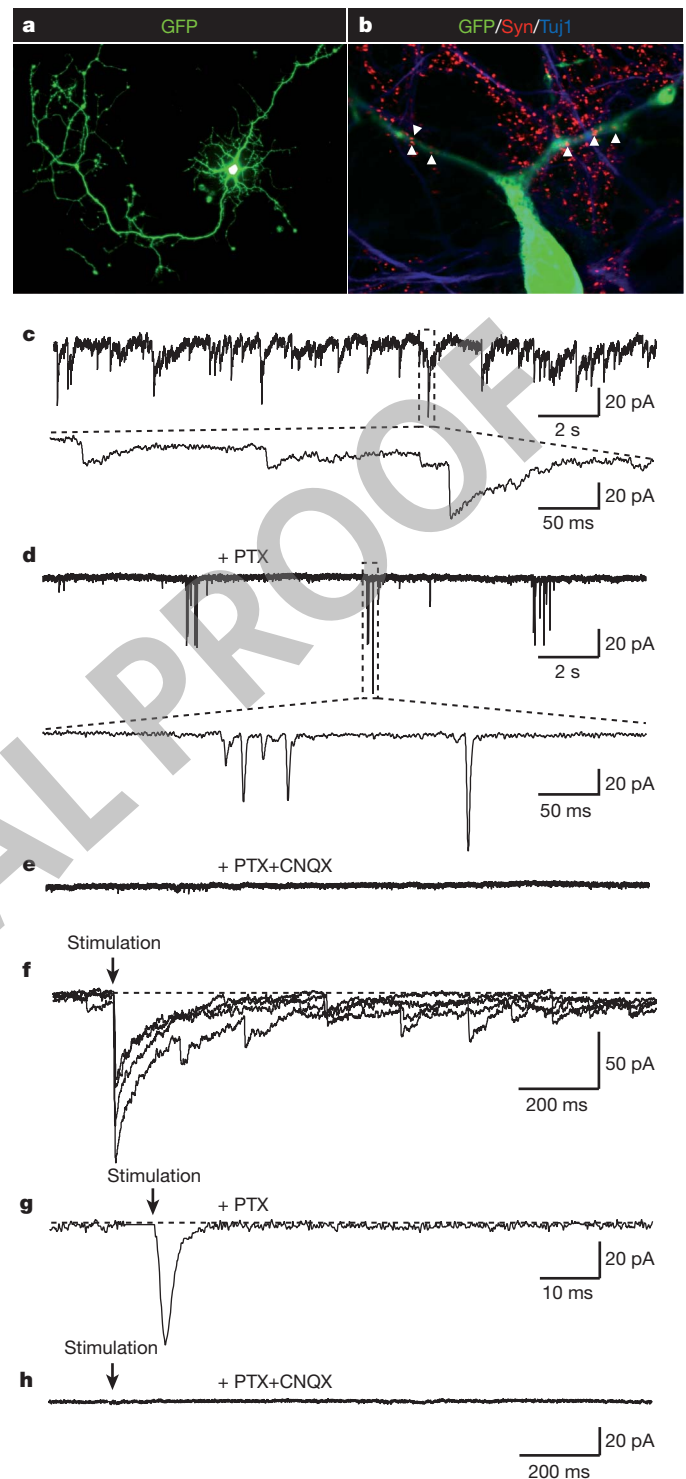


Figure 4 | Synaptic responses of HFF-iN cells. **a**, An HFF-iN cell expressing EGFP co-cultured with mouse cortical neurons at day 35 after doxycycline treatment. **b**, Synapsin-positive puncta co-localize with neurites extending from HFF-iN cells (arrow heads). **c**, Thirty-five days after doxycycline treatment, spontaneous PSCs were recorded in HFF-iN cells. **d**, Slow responses could be blocked by picrotoxin (PTX). The insert shows the fast kinetics of the responses. **e**, In the presence of PTX and CNQX (both $50 \mu\text{M}$), no spontaneous activities were observed. **f**, Evoked postsynaptic responses. Four traces were superimposed. **g**, In the presence of PTX, electric stimulation evoked fast-kinetic excitatory PSCs (EPSCs). **h**, No evoked synaptic responses were observed in the presence of PTX and CNQX. Scale bars, 100 μm (**a**); 10 μm (**b**).

pan-neuronal markers (Fig. 3e–g). Intriguingly, the efficiencies of iN cell generation from fetal and postnatal fibroblasts were similar (2–4% of cells plated; Fig. 3a). Single-cell gene-expression profiling of iN cells revealed that 46 of 51 HPF iN cells co-expressed pan-neuronal and synaptic markers 42 days after infection; the majority of the HPF iN cells (37 of 51) seem to be glutamatergic neurons. Immunofluorescence analysis showed that 6 weeks after infection, $81 \pm 17\%$ of iN cells expressed Tbr1 above the levels of fibroblasts, and $15.2 \pm 6.6\%$ were peripherin-positive (Supplementary Fig. 13). Unlike fetal fibroblasts, most postnatal fibroblasts showed weak but specific Tbr1 staining (Supplementary Fig. 13a–d). Electrophysiological recordings demonstrated the presence of regenerative action potentials as well as voltage-dependent channel activities in the majority of cells analysed from two lines (for example, 17 of 18 cells from line HPF-B, Fig. 3h, i and Supplementary Table 1). Furthermore, iN cells with active membrane properties could be generated from dermal fibroblasts derived from an 11-year-old human subject (Supplementary Fig. 14).

Finally, we determined whether human iN cells can express functional neurotransmitter receptors and form functional synapses. Application of either GABA (γ -aminobutyric acid) or L-glutamate to HPF-iN cells induced current responses that could be blocked by picrotoxin and CNQX, respectively (Supplementary Fig. 12e, f). We then dissociated HFFs 4–7 days after infection with the BAMN factors and EGFP and plated them onto previously established mouse cortical neuronal cultures. These co-cultures were maintained for up to 5 weeks thereafter. HFF-iN cells were identified by the EGFP expression (Fig. 4a). Whole-cell recordings after 2–3 weeks of co-culture showed no synaptic activity ($n = 20$) but after 4–5 weeks approximately half of human iN cells recorded showed spontaneous post-synaptic currents (PSCs) with variable kinetics ($n = 21$, Fig. 4c and Supplementary Table 1). Immunostaining with synapsin antibodies confirmed the presence of scattered synaptic puncta on the dendrites of EGFP-positive cells (Fig. 4b). When the GABA_A receptor inhibitor picrotoxin was applied, the majority of the spontaneous PSCs were blocked, demonstrating that they were inhibitory (IPSCs) (Fig. 4d). In the presence of picrotoxin, bursting of spontaneous excitatory post-synaptic currents (EPSCs) was shown and could be blocked by the AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptor blocker CNQX (Fig. 4e). Focal stimulation evoked both IPSCs and EPSCs that could be blocked by picrotoxin and CNQX (Fig. 4f–h). Importantly, PSCs could also be recorded from HPF-iN cells co-cultured with mouse cortical neurons 4 weeks after infection (Supplementary Fig. 15). These results demonstrate that fetal and postnatal fibroblast-derived iN cells could form functional synapses and integrated into pre-existing neuronal networks.

In this report, we have identified a combination of transcription factors that are capable of converting human fibroblasts directly into neurons. Like mouse iN cells¹³ and neurons derived from ES cells^{19–21} and iPS cells^{22,23}, the human iN cells seem relatively immature, as indicated by their slightly depolarized membrane potentials and the relatively low-amplitude synaptic responses. Compared to mouse iN cells, human iN cells required longer culture periods to develop synaptic activity. Future studies will be necessary to thoroughly optimize conditions for human iN cell generation and maturation, which would facilitate applications of this method for the study of human neuronal development and disease.

METHODS SUMMARY

Cell culture. H9 human ES cells (WiCell Research Resources) and iPS cells were expanded in mTeSR1 (Stem Cell Technologies) and passaged as clumps or single cells²⁴. Primary HFFs were isolated from the distal half of the limbs of GW8–10 fetuses obtained from Advanced Bioscience Resources Inc. Primary HPFs were established from foreskin. Primary mouse cortical cultures and glial monolayer cultures were established as described previously¹³.

Lentiviral infections. Lentiviral production and fibroblast infections were performed as described previously¹³. Primary fibroblasts and pluripotent stem cells

were infected with concentrated lentivirus and treated with doxycycline ($2 \mu\text{g} \mu\text{l}^{-1}$) 16–24 h later.

Electrophysiology and expression analysis. Cells were analysed by immunofluorescence and electrophysiology as described elsewhere^{13,25}. Single-cell gene expression profiling was performed using the Fluidigm Biomark dynamic array^{18,26} according to the manufacturer's instructions.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 November 2010; accepted 18 May 2011.

Published online 26 May 2011.

- Blau, H. M. *et al.* Plasticity of the differentiated state. *Science* **230**, 758–766 (1985).
- Gurdon, J. B. From nuclear transfer to nuclear reprogramming: the reversal of cell differentiation. *Annu. Rev. Cell Dev. Biol.* **22**, 1–22 (2006).
- Heins, N. *et al.* Glial cells generate neurons: the role of the transcription factor Pax6. *Nature Neurosci.* **5**, 308–315 (2002).
- Ieda, M. *et al.* Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**, 375–386 (2010).
- Shen, C.-N., Slack, J. M. W. & Tosh, D. Molecular basis of transdifferentiation of pancreas to liver. *Nature Cell Biol.* **2**, 879–887 (2000).
- Tada, M., Takahama, Y., Abe, K., Nakatsuji, N. & Tada, T. Nuclear reprogramming of somatic cells by *in vitro* hybridization with ES cells. *Curr. Biol.* **11**, 1553–1558 (2001).
- Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
- Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810–813 (1997).
- Xie, H., Ye, M., Feng, R. & Graf, T. Stepwise reprogramming of B cells into macrophages. *Cell* **117**, 663–676 (2004).
- Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D. A. *In vivo* reprogramming of adult pancreatic exocrine cells to β -cells. *Nature* **455**, 627–632 (2008).
- Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462**, 587–594 (2009).
- Zhou, Q. & Melton, D. A. Extreme makeover: converting one cell into another. *Cell Stem Cell* **3**, 382–388 (2008).
- Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
- Hansen, D. V., Lui, J. H., Parker, P. R. & Kriegstein, A. R. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010).
- Kriegstein, A., Noctor, S. & Martinez-Cerdeno, V. Patterns of neural stem and progenitor cell division may underlie evolutionary cortical expansion. *Nature Rev. Neurosci.* **7**, 883–890 (2006).
- Zhang, X. *et al.* Pax6 is a human neuroectoderm cell fate determinant. *Cell Stem Cell* **7**, 90–100 (2010).
- Bottenstein, J. E. & Sato, G. H. Growth of a rat neuroblastoma cell line in serum-free supplemented medium. *Proc. Natl Acad. Sci. USA* **76**, 514–517 (1979).
- Guo, G. *et al.* Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* **18**, 675–685 (2010).
- Johnson, M. A., Weick, J. P., Pearce, R. A. & Zhang, S. C. Functional neural development from human embryonic stem cells: accelerated synaptic activity via astrocyte coculture. *J. Neurosci.* **27**, 3069–3077 (2007).
- Wu, H. *et al.* Integrative genomic and functional analyses reveal neuronal subtype differentiation bias in human embryonic stem cell lines. *Proc. Natl Acad. Sci. USA* **104**, 13821–13826 (2007).
- Koch, P., Opitz, T., Steinbeck, J. A., Ladewig, J. & Brustle, O. A rosette-type, self-renewing human ES cell-derived neural stem cell with potential for *in vitro* instruction and synaptic integration. *Proc. Natl Acad. Sci. USA* **106**, 3225–3230 (2009).
- Marchetto, M. C. *et al.* A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell* **143**, 527–539 (2010).
- Hu, B. Y. *et al.* Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. *Proc. Natl Acad. Sci. USA* **107**, 4335–4340 (2010).
- Xu, Y. *et al.* Revealing a core signaling regulatory mechanism for pluripotent stem cell survival and self-renewal by small molecules. *Proc. Natl Acad. Sci. USA* **107**, 8129–8134 (2010).
- Maximov, A., Pang, Z. P., Tervo, D. G. & Sudhof, T. C. Monitoring synaptic transmission in primary neuronal cultures using local extracellular stimulation. *J. Neurosci. Methods* **161**, 75–87 (2007).
- Wong, C. C. *et al.* Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage. *Nature Biotechnol.* **28**, 1115–1121 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank Y. Kokubu for technical assistance in molecular cloning and Y. Zhang for assistance in iPS cell induced neuron culture. We also thank Y. Sun for providing the microRNAs expression lentiviral vectors and S. Majumder for the REST-VP16 construct. This work was enabled by start-up funds from the Institute for Stem Cell Biology and Regenerative Medicine at Stanford (M.W.), the Ellison Medical Foundation (M.W.), the Stinehard-Reed Foundation (M.W.), the Donald E. and Delia B. Baxter Foundation (M.W.), the NIH grants 1R01MH092931

(M.W. and T.C.S.) and RC4 NS073015 (M.W.), and a Robertson Investigator Award from the New York Stem Cell Foundation. Z.P.P. is supported by 2008 and 2010 NARSAD Young Investigator Awards. T.V. is supported by the Ruth and Robert Halperin Stanford Graduate Fellowship. A.C. is supported by the AXA research fund and D.R.F. is supported by BioX Undergraduate Fellowship.

Author Contributions Z.P.P., N.Y., T.V., A.O., T.C.S. and M.W. designed the experiments and analysed the data. D.R.F. and T.Q.Y. helped with lentiviral production. A.C., V.S. and

S.M. helped to provide experimental material and helped with the analyses. Z.P.P., N.Y., T.V., T.C.S. and M.W. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to (wernig@stanford.edu).

NOT FINAL PROOF

METHODS

Cell culture. H9 human ES cells (WiCell Research Resources) were expanded in mTeSR1 (Stem Cell Technologies). Induced pluripotent stem cells were generated as described elsewhere²⁷. The day before infection, cells were treated with Accutase and seeded as single cells in 3.5-cm tissue culture dishes on Matrigel in mTeSR1 containing 2 μ M thiazovivin (Bio Vision)²⁴. To inhibit the growth of uninfected ES cells and select for post-mitotic neurons, we added 4 μ M cytosine β -D-arabinofuranoside (Ara-C) to the media 48 h after doxycycline addition. Primary human fetal fibroblasts were isolated from the distal half of the limbs of 8–10-week-old fetuses (Advanced Bioscience Resources Inc.). The tissue was dissociated after trypsin digestion and plated in MEF media (DMEM high glucose, calf serum, sodium pyruvate, non-essential amino acids, penicillin/streptomycin and β -mercaptoethanol). Primary human postnatal fibroblasts (HPFs) were established from dissociated foreskin tissue derived from 1–3-day-old newborns. Before being used for experiments, primary fibroblast cells were passaged at least three times. Primary mouse cortical cultures and glial monolayer cultures were established as described previously¹³. To maintain the iN cell culture, cells were either grown in N3 medium (DMEM/F2 (Invitrogen), apotransferrin (100 μ g ml⁻¹), insulin (5 μ g ml⁻¹), sodium selenite (30 nM), progesterone (20 nM), putrescine (100 nM), penicillin/streptomycin) supplemented with neurotrophic factors including brain-derived neurotrophic factor, glial-cell-derived neurotrophic factor, neurotrophin-3 and ciliary neurotrophic factor (R&D systems), or dissociated using papain and replated onto previously established monolayer culture of primary mouse glia or primary neurons from mouse cortex in neuronal growth medium (MEM (Gibco) supplemented with B27 (Gibco), glucose (5 mg ml⁻¹), transferring (10 μ g ml⁻¹), 5% fetal bovine serum and Ara-C (2 μ M, Sigma)^{13,25}.

Virus infections. Lentiviral production and fibroblast infections were performed as described previously¹³. Briefly, primary human fetal or postnatal fibroblasts were plated and infected with concentrated lentiviral particles and polybrene (8 μ g μ l⁻¹) in fresh MEF medium. Viral medium was removed after 16–24 h and replaced with MEF medium containing doxycycline (2 μ g μ l⁻¹). After 24–48 h, medium was changed to N3 medium containing doxycycline (2 μ g μ l⁻¹). For human ES cell infections, H9 human embryonic stem cells were switched into N3 medium containing polybrene (2 μ g μ l⁻¹) 24 h after re-plating, and concentrated lentiviral particles were added. After 16–24 h, cultures were switched to N3 medium containing doxycycline (2 μ g μ l⁻¹) and changed daily before dissociation. Forty-eight hours after the initial addition of doxycycline, Ara-C (4 μ g μ l⁻¹) was added to the medium to inhibit proliferation of uninfected ES cells until analysis 6 days after the addition of doxycycline. All chemicals were purchased from Sigma-Aldrich if not otherwise specified.

Single-cell gene expression analysis (Fluidigm dynamic array). Single-cell gene expression profiling was performed using the Fluidigm Biomark dynamic array according to the manufacturer's protocol^{18,26}. Briefly, single cells growing on culture dishes after 5 or 6 weeks of transduction were collected by aspiration into patch electrodes and ejected into 2 \times cellsdirect buffer (Invitrogen), flash-frozen and kept at -80°C until further processing. Thawed cells were subject to target-specific reverse-transcription and 18 cycles of PCR pre-amplification with a mix of primers specific to the target genes (STA). STA products were then processed for real-time PCR analysis on Biomark 48:48 Dynamic Array integrated fluidic circuits (Fluidigm). To ensure the specificity of the amplification, titrations of total human brain RNA were included in each experiment, and only primers that demonstrated a linear amplification were analysed. Furthermore, melting curves of the PCR products were compared between the single cells and the control RNA to ensure the specificity of the PCR products.

Electrophysiology. Action potentials were recorded with current-clamp whole-cell configuration. The pipette solution for current-clamp experiments contained (in mM): 123 K-gluconate, 10 KCl, 1 MgCl₂, 10 HEPES, 1 EGTA, 0.1 CaCl₂, 1 K₂ATP, 0.2 Na₄GTP and 4 glucose, pH adjusted to 7.2 with KOH. Membrane potentials were kept around -65 to -70 mV, and step currents were injected to elicit action potentials. For whole-cell voltage-dependent current recordings, the same internal solution as aforementioned was used. For synaptic functional evaluation, the internal solution contained (in mM): 135 CsCl, 10 HEPES, 1 EGTA, 4 Mg-ATP, 0.4 Na₄GTP, and 10 QX-314, pH 7.4. The bath solution contained (in mM): 140 NaCl, 5 KCl, 2 CaCl₂, 2 MgCl₂, 10 HEPES, and 10 glucose, pH 7.4. Synaptic responses were measured as described previously^{13,25}. Stimulus artefacts for evoked synaptic responses were removed for graphic representation. Electrophysiological data are presented as mean \pm s.e.m. All statistical comparisons were made using Student's *t*-test.

Immunofluorescence and RT-PCR. For immunofluorescence experiments, cells were fixed in 4% paraformaldehyde in PBS for 10 min at room temperature. After fixation, cells were incubated in 0.2% Triton X-100 in PBS for 5 min at room temperature. After washing twice with PBS, cells were blocked in a solution of PBS containing 4% BSA and 1% Cosmic calf serum (CCS) for 30 min at room temperature. Primary and secondary antibodies were applied for 1 h and 30 min, respectively. The following antibodies were used for our analysis: rabbit anti-Tuj1 (Covance, 1:1,000), mouse anti-Tuj1 (Covance, 1:1,000), mouse anti-MAP2 (Sigma, 1:500), mouse anti-NeuN (Millipore, 1:200), mouse anti-neurofilament (Developmental Studies Hybridoma Bank (DSHB), 2H3a, 1:1,000), rabbit anti-synapsin (E028, 1:1,000), anti-synaptotagmin (Synaptic systems, 41.1, 1:2,000), guinea pig anti-vGLUT1 (Millipore, 1:2,000), mouse anti-GAD6 (DSHB, 1:500), rabbit anti-Tbr1 (Abcam, 1:200), mouse anti-peripherin (Sigma, 1:100), sheep anti-tyrosine hydroxylase (Pel-Freez, 1:500), rabbit anti-GEAP (DAKO, 1:4,000), mouse anti-Sox2 (R&D Systems, 1:50), goat anti-Brn2 (clone C-20, Santa Cruz Biotechnology, 1:100), rabbit anti-Ascl1 (Abcam, 1:200), mouse anti-BrdU (Becton Dickinson, 1:50), mouse anti-LU5 (Abcam, 1:200), goat anti-Sox10 (Santa Cruz Biotechnology, 1:40). Fluorescein isothiocyanate (FITC)-, and Cy3-conjugated secondary antibodies were obtained from Jackson ImmunoResearch. Alexa-488-, Alexa-546- and Alexa-633-conjugated secondary antibodies were obtained from Invitrogen. 4',6-Diamidino-2-phenylindole (DAPI) was from Sigma (1:10,000). For RT-PCR analysis, RNA was isolated using an RNAqueous Kit (Applied Biosystems) following the manufacturer's instructions, treated with DNase (Applied Biosystems) and reverse-transcribed with Superscript III (Invitrogen).

Efficiency calculation. The following method was used to calculate the efficiency of neuronal induction. The total number of Tuj1-positive cells with a neuronal morphology, defined as cells having a circular, three-dimensional appearance that extend a thin process at least three times longer than their cell body, were quantified at indicated time points. We determined this number in at least 15 randomly selected $\times 20$ visual fields on a Zeiss Axio Observer microscope. The average and standard deviation per field was determined and then used to extrapolate the total number of iN cells present in the entire dish based the known surface areas of a $\times 20$ visual field and the respective culture dish. We then divided this number by the number of cells plated before infection to get the percentage of the starting population of cells that adopted neuron-like characteristics. In multiple independent experiments we verified that this extrapolation method yields cell numbers very similar to those measured by a haemocytometer at the time of plating. Data are presented as mean \pm s.d.

27. Somers, A. *et al.* Generation of transgene-free lung disease-specific human induced pluripotent stem cells using a single excisable lentiviral stem cell cassette. *Stem Cells*, **28**, 1728–1740 (2010).